ABSTRACT

Title of Dissertation: PSYCHOMETRIC ASSESSMENT OF TWO NEW
                             SELF-RATING DEPRESSION SCALES: THE
                             SCHIRALDI DEPRESSION CHECK-UP AND THE
                             CORREA-BARRICK DEPRESSION SCALE

Christina Barrett Barrick, Doctor of Philosophy, 1994

Dissertation Directed by: Harvey E. Clearwater, Ed.D.
                             Associate Professor
                             College of Health and Human
                             Performance, Department of Health
                             Education

Depression is a prevalent condition that is responsive to
treatment. Efforts to screen and educate the public on depression
are beneficial. The purpose of this investigation was to assess
the psychometric properties of two new self-rating depression
scales, the Schiraldi Depression Check-up (DC) and the Correa-
Barrick Depression Scale (CBDS), based upon classical test theory
and comparisons to published scales: the Beck Depression Inventory
(BDI) and the Inventory for Depressive Symptomatology, Self-Report
(IDS-SR).

The study was conducted on a total of 387 participants.
There were two convenience samples used. Sample I was composed of
337 faculty and staff from a metropolitan comprehensive university.
A subset of Sample I was composed of 203 faculty and staff. Sample
II included 50 outpatients diagnosed with depression under the
treatment of a board certified psychiatrist.

Correlation coefficients for the DC and CBDS with the BDI were $r = .75$ and $r = .71$, respectively. Cronbach alpha coefficients for the DC and CBDS in the patient sample were $r = .95$ and $r = .96$, respectively. Correlation coefficients for the IDS-SR with the DC and the CBDS in the patient sample were $r = .85$ and $r = .81$, respectively. Two-week test-retest correlation coefficients in the university sample for the DC and CBDS were $r = .81$ and $r = .70$, respectively. Factor analysis for the DC revealed a three-factor structure: "Cognitive-Emotional Disturbance," "Psychophysiological Symptoms," and "Physiological Symptoms." Factor analyses for the CBDS revealed a four-factor structure: "Cognitive-Emotional Disturbance," "General Outlook," "Physiological Symptoms," and "Sensory/Perceptual Disturbance." Discriminant analysis did not support the Depression Check-up or the Correa-Barrick Depression Scale in discriminating between the university and patient samples.

There were several conclusions from this study. Findings provided preliminary evidence for the validity and reliability of the Depression Check-up and the Correa-Barrick Depression Scale in measuring depression in an adult population. The findings that sensory-perceptual disturbance may be an additional variable in depression and that a single-factor structure emerged for "emotional-cognitive disturbance" was discussed as well as implications for health education theory, practice, and research.

PSYCHOMETRIC ASSESSMENT OF TWO NEW SELF-RATING DEPRESSION SCALES:

THE SCHIRALDI DEPRESSION CHECK-UP AND

THE CORREA-BARRICK DEPRESSION SCALE

by

Christina Barrett Barrick

Dissertation submitted to the Faculty of the Graduate School
of The University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1994

Advisory Committee:

    Professor Glen G. Gilbert
    Associate Professor Kenneth H. Beck
    Associate Professor Harvey E. Clearwater, Chairman/Advisor
    Associate Professor William D. Schafer
    Instructor Glenn R. Schiraldi

DEDICATION

In memory of my mother, Audrey C. Barrett

ACKNOWLEDGEMENT

My advisor, Dr. Harvey Clearwater, and Committee members

My husband, Brian

My family, friends, faculty colleagues and mentors

Pat Bourgeois, my editor and typist

and

Elsa Correa, M.D., and Joan Wagner

TABLE OF CONTENTS

II. continued

LIST OF TABLES

CHAPTER I

INTRODUCTION

Depression is a prevalent condition that affects more than
six million Americans each year (Weissman et al., 1988). Zung
(1990) reported the cost of depression due to lost productivity to
be estimated at $14.2 billion. In addition, depressed persons
utilize health care services more frequently than non-depressed
persons. It has been reported that individuals with depression
make three times as many total health care visits as patients
without psychiatric illness (Regier et al., 1988; Katon & Sullivan,
1990; Shapiro et al., 1984). Depressed persons also make more
ambulatory physician visits, place more telephone calls, and have
more medical evaluations (Katon & Sullivan, 1990).

In order to promote awareness of depression, the National
Institute of Mental Health (NIMH) launched a comprehensive
Depression Awareness, Recognition, and Treatment Program (Regier et
al., 1988). Despite efforts, problems with the recognition and
management of depression still exist in the primary care setting
(Zung, 1990).

While the reasons for underrecognition of depression are
varied, patients in primary care settings are likely to report

1

somatic complaints and symptoms suggesting organic diseases (Magruder-Habib, Zung & Feussner, 1990), as depression can be masked by anxiety or physical complaints such as backache, headache, or gastrointestinal distress. About 66% of undiagnosed depressed patients make more than six visits a year to primary care physicians for treatment of somatic complaints (Katon & Sullivan, 1987). When depression is detected, inappropriate medications such as anxiolytics, analgesics, and sedatives are often given. These drugs can mask the depression or, worse, deepen the depression. An NIMH study found that as many as half of all depressed patients were treated with anxiolytics instead of antidepressants (Keller et al., 1982; Keller, Lavori & Klerman, 1986).

Depression also causes higher morbidity than other chronic conditions. The Medical Outcomes Study (Wells et al., 1989) compared the overall state of health of depressed patients with those with other chronic illnesses such as heart disease, hypertension, diabetes, arthritis, and back pains. They found that the only chronic conditions that cause levels of disability comparable with those seen with depressed persons were advanced coronary artery disease and angina. Specifically, depressed persons spent more days in bed than did persons with other chronic illnesses and they complained of more bodily pain than any other group except the groups of persons with arthritis (Wells et al., 1989).

Mortality among depressed persons is also noteworthy. Not only is mortality from suicide an expected complication of depression (Johnson, Weissman & Klerman, 1992), other studies found a distressing association between depression and increased cardiovascular mortality. In a prospective study, Dreyfuss, Dashber and Assael (1969) found an association between depression and myocardial infarction and concluded that "the depression could not be regarded as a reaction to the impact of the myocardial infarction" (p. 80). Furthermore, Tsuang, Woolson and Fleming (1980) reported that survival rates for depressives were shortened compared to the general population. The argument that antidepressant medication, specifically tricyclics, accounted for the increased mortality among depressed persons was not supported by research studies which controlled for medication side effects. Mondimore (1990) and Weeke, Juel and Vaeth (1987) reported that patients with affective disorders had a slightly significantly higher risk for mortality from cardiovascular disease relative to the general population even after controlling for medication.

Underrecognition of depression in the health care system has serious costs. Medical patients with psychiatric conditions make more office visits to their primary care physician, are frequently referred to specialists for additional medical workups, are subjected to unnecessary, potentially harmful, and costly diagnostic tests, and are subjected to the consequences of improper treatment (Zung, 1990). Since depression may present an obscure

clinical picture, physicians, understandably, order diagnostic tests for depressed persons to protect against malpractice.

Rationale for a New Scale

There are compelling reasons for valid and reliable scales for identifying depression in primary care settings. The case for use of depression scales in such settings was strengthened in a study by Magruder-Habib, Zung, and Feussner (1990) who found significant differences in primary care physician recognition and treatment of depression when a depression screening scale was used. For busy primary care physicians and psychiatrists, self-rating scales can provide a useful alternative to the more time-consuming interviewer rating scales (Zung, 1990). Hence, several reasons for developing a new depression self-rating scale became apparent.

First, problems exist with current, published depression self-rating scales. For example, although the Beck Depression Inventory and the Zung Self-rating Depression Scale are popular scales with known psychometric properties, both lack some of the current diagnostic criteria for depression found in the Revised Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association [APA], 1987). (The specific items that each scale is missing will be described in Chapter II).

Second, many depression scales contain items that reflect the medical model; that is, items are symptom-oriented and do not always capture qualitative aspects or feelings of the depression.

Third, none of the scales reviewed by the investigator contained items on decreased sensory awareness such as impaired taste and lack of color. Enhanced sensory awareness in mania was reported in the literature (Goodwin & Jamison, 1990) but not among depressed patients. Nonetheless, some preliminary evidence for adding these items to enhance the validity of a depression scale were found in the literature (to be discussed in Chapter II) and from interviews with psychiatrists (described in the "procedure" section).

Purpose of the Study

The purpose of this study was twofold: to develop the Correa-Barrick Depression Scale and to establish initial validity and reliability of two new self-rating depression scales: the Depression Check-up (Schiraldi, 1987; see Appendix A) and the Correa-Barrick Depression Scale (Appendix B). Using survey research design, data from two different population samples were analyzed.

Population Sample I consisted of a representation of faculty and staff at a metropolitan comprehensive university, while Sample II was made up of patients who met the current diagnostic criteria for a major depression and were being treated by a board certified psychiatrist.

Other known, published instruments were used in the study to establish validity. The Beck Depression Inventory was used in the

pilot study and the Inventory for Depressive Symptomatology, Self-report (IDS-SR) was used in both the pilot and the main studies.

There were three phases to this reliability and validity study: Phase I, qualitative assessment of the Correa-Barrick Depression Scale using focus group interviews and expert panel review; Phase II, pretest of the Correa-Barrick Depression Scale and the Depression Check-up in a pilot study, and Phase III, the final, main study. Each phase will be described in Chapter III.

A self-rating depression scale should possess certain criteria. Therefore, based on a review of the literature on scale development and depression assessment, as well as the investigator's beliefs about how the scales should be utilized, criteria were developed by the investigator and included the following components or attributes: (a) self-rating; (b) brevity (that is, 10 minutes or less to complete); (c) paper and pencil administration; (d) ease of scoring and interpretation by an interdisciplinary health care team; (e) items reflecting all of the diagnostic criteria for depression as defined in the revised Diagnostic and Statistical Manual of Mental Disorders (APA, 1987); (f) items that can be used in health education for screening persons for depression and/or program evaluation; (g) items that can be used in primary care settings for screening persons for depression, and (h) sensitivity (that is, can be used for evaluating patient outcomes to stress management, psychotherapy and/or medication effectiveness).

Research Questions and Hypotheses for the Pilot Study

The main purpose of the pilot test was to pretest the Depression Check-up and the Correa-Barrick Depression Scale by gathering preliminary reliability and validity data about the two instruments. The Depression Check-up and the Correa-Barrick Depression Scale also needed to be piloted for several other reasons: (a) to determine clarity of the instructions and item wording; (b) to evaluate formats of the scales; (c) to assess any respondent difficulty in following directions or completing the questionnaires; (d) to assess scale items using an item analysis; (e) to evaluate potential problems in conducting the psychometric evaluation study, and (f) to evaluate potential problems with the two-week test-retest method in the university sample in the key areas of response rate, clarity of instructions and directions, and problems with participant self-coding. Data from the pilot study were used to revise the survey prior to the main study.

Research questions for the pilot study were:

1. Will the Depression Check-up and the Correa-Barrick Depression Scale possess convergent validity with the Beck Depression Inventory?

2. Will the Depression Check-up and the Correa-Barrick Depression Scale possess convergent validity with the Inventory for Depressive Symptomatology?

3. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate concurrent validity?

4. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability with internal consistency?

5. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability using the coefficient of stability from a single-test administration?

These research questions led to the following hypotheses:

Hypothesis 1.

A. The Depression Check-up will positively correlate with the Beck Depression Inventory.

B. The Correa-Barrick Depression Scale will positively correlate with the Beck Depression Inventory.

Hypothesis 2.

A. The Depression Check-up will positively correlate with the Inventory for Depressive Symptomatology, Self-report.

B. The Correa-Barrick Depression Scale will positively correlate with the Inventory for Depressive Symptomatology, Self-report.

Hypothesis 3.

A. The total depression score from the Depression Check-up will correlate with the following scale variable from the eighth question of the Depression Check-up: "Sad (gloomy, discouraged, blue, numb, empty, or like you just don't care)."

B. The total depression score from the Correa-Barrick Depression Scale will correlate with the following scale variable from Question 21 of the Correa-Barrick Depression Scale: "I feel depressed (sad, blue, and gloomy)."

Hypothesis 4.

A. The Depression Check-up will demonstrate a positive Cronbach's alpha coefficient.

B. The Correa-Barrick Depression Scale will demonstrate a positive Cronbach's alpha coefficient.

Hypothesis 5.

A. The Depression Check-up will demonstrate a positive correlation coefficient on the split-half procedure (odd/even).

B. The Correa-Barrick Depression Scale will demonstrate a positive correlation on the split-half procedure (odd/even).

Research Questions and Hypotheses of the Main Study

There were six research questions and six research hypotheses for the main study. The research questions were:

1. Will the Depression Check-up and the Correa-Barrick Depression Scale possess convergent validity?

2. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate concurrent validity?

3. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability based upon internal consistency?

4. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability over time?

5. Will the Depression Check-up and the Correa-Barrick Scale demonstrate construct validity?

6. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate discrimination?

These research questions led to the following hypotheses:

Hypothesis 1.

   A. The Depression Check-up will positively correlate with the Inventory for Depressive Symptomatology, Self-report.

   B. The Correa-Barrick Depression Scale will positively correlate with the Inventory for Depressive Symptomatology, Self-report.

Hypothesis 2.

   A. The total depression score from the Depression Check-up will positively correlate with the following survey variables: family history, personal depression history, taking medications for depression, current depression, and question II-A of the Depression Check-up: " Sad (gloomy, discouraged, blue, numb, empty, or like you just don't care)."

   B. The total depression score from the Correa-Barrick Depression Scale will positively correlate with the following

survey variables: family history, personal depression history, taking antidepressants, current depression, and Question 13 of the Correa-Barrick Depression Scale: "I feel depressed (sad, blue, and gloomy)."

Hypothesis 3.

A. The Depression Check-up will demonstrate a positive Cronbach's alpha.

B. The Correa-Barrick Depression Scale will demonstrate a positive Cronbach's alpha.

Hypothesis 4.

A. Depression scores on the Depression Check-up will positively correlate with the depression scores on the Depression Check-up when administered two weeks or 14 days later.

B. Depression scores on the Correa-Barrick Depression Scale will positively correlate with the depression scores on the Depression Check-up when administered two weeks or 14 days later.

Hypothesis 5.

A. The Depression Check-up will cluster around three dimensions from factor analysis: cognitive impairment, physiological symptoms, and emotional distress.

B. The Correa-Barrick Depression Scale will cluster around four dimensions from factor analysis: cognitive impairment, physiological symptoms, emotional distress, and decreased sensory awareness.

Hypothesis 6.

    A. The Depression Check-up will discriminate between the university sample and the patient sample.

    B. The Correa-Barrick Depression Scale will discriminate between the university sample and the patient sample.

In order to avoid confusion about the expected direction for individual statistical tests, each hypothesis is stated as the expected research outcome but not necessarily in the null format unless no differences were expected.

Limitations of the Study

There were limitations to this study.

1. A convenience sample was used which limited the study's generalizability to the population characteristics of the faculty and staff sample and the sample of patients diagnosed with depression who were being treated in a private practice.

2. The instruments developed in this study were atheoretical. Although theories from research studies exist about the cause or causes of depression, the cause remains unknown. At present, a theory about the cause of depression would need to be extraordinarily complex involving biological, psychosocial, and environmental factors and would be beyond the scope of this study.

3. Results of this study should be considered only as an initial step in establishing the validity and reliability of the

two depression scales. Further replications in different populations would be preferable before implementing the instrument.

4. Cut-off scoring of the self-rating scales for depression was based on the norms scores from the population used; hence, the norms may change with a different population. One canon for interpretation of the scores is that multiple measures should be used before a person is diagnosed with major depression. Multiple measures should include but not be limited to physical and laboratory examinations by a physician, clinical interview and observations, and persistence of the depressed mood beyond two weeks.

5. The Beck Depression Inventory was used for establishing criterion (convergent) validity but only for the pilot sample. Due to the method of administration (mailed surveys), security of the test could not be guaranteed; therefore, permission could not be obtained for reproduction of the Beck Depression Inventory for administration to a larger sample.

6. Only one definition of depression was used for this study, the rationale for which will be explained. To enhance agreement among clinicians and investigators, the American Psychiatric Association developed the Diagnostic and Statistical Manual of Mental Disorders from a consensus of a panel of experts. In existence since 1952, it was the first official manual of mental disorders to contain a glossary of descriptions of diagnostic categories. Since then it has undergone extensive revision. The

latest manual (DSM-III-R) was revised in 1987 because new research data had emerged and the diagnostic criteria needed to be reviewed for accuracy. The DSM-III-R is widely accepted in the United States by clinicians and researchers, is used internationally, and has been cited extensively (over 2,000 research articles) in the scientific literature. The DSM-III-R is being updated to the DSM-IV version.

Authors of the DSM-III-R assumed a certain philosophy about major depression: that it is a mental disorder involving complex behavioral, psychological, and physiological systems and can impair functional capacity and cause distress or an increased risk of suffering pain, disability, and death.

There are variations in clinical features of depression based on age, especially in children, adolescents, and the elderly, which were noted in the DSM-III-R. This may be an issue in using the DSM-III-R in populations other than adults.

Spitzer and Williams (cited in APA, 1987) reported some limitations to the DSM-III-R. First, it provides descriptive information about major depression as a diagnostic category but does not furnish information about the etiology of depression. In this respect, the DSM-III-R does not have a basis in a theoretical framework (APA, 1987). Secondly, the DSM-III-R provides only behavioral observations and symptomatology; it does not capture the phenomenological, qualitative experiences of the person suffering from depression. In other words, the DSM-III-R is oriented to the

medical model since it is "symptom-oriented". For example, Criterion #1 refers to depressed mood but does not include qualifying descriptions for depressed mood. Although the DSM-III-R is not conceptually flawless, it does provide some measurable criteria for both the researcher and/or clinician and remains the current standard in mental health classification.

Definition of Terms

For purposes of this study, the following definitions were used.

Cognitive disturbance. Diminished capacity to realistically appraise or "assign meanings to the world as it unfolds before us" (Everly, 1989, p. 25).

Depression. A syndrome characterized by a number of features in addition to the depressed mood which is generally considered to be a necessary but not sufficient indicator for the diagnosis (Feighner & Boyer, 1991). According to the APA (1987), although other names are given to "depression" (major depression, affective disorder, mood disorder, and unipolar/bipolar depression), it still refers to the same syndrome. Depression is operationally defined by the current Diagnostic and Statistical Manual (DSM-III-R) of the American Psychiatric Association (APA, 1987, pp. 222-223) as having at least five of the following nine symptoms present during the same two-week period, representing a change from previous

functioning, with one of the symptoms being either depressed mood
or loss of interest or pleasure.

1. depressed mood (or can be irritable mood in
children and adolescents) most of the day, nearly every day, as
indicated either by subjective account or observation by others;

2. markedly diminished interest or pleasure in all or
almost all activities most of the day, nearly every day.

3. significant weight loss or weight gain when not
dieting (e.g., more than 5% of body weight in a month) or decrease
or increase in appetite nearly every day (in children, consider
failure to make expected weight gains);

4. insomnia or hypersomnia nearly every day;

5. psychomotor agitation or retardation nearly every
day;

6. fatigue or loss of energy nearly every day;

7. feelings of worthlessness or excessive or
inappropriate guilt;

8. diminished ability to think or concentrate or
indecisiveness nearly every day;

9. recurrent thoughts of death (not just fear of
dying), recurrent suicidal ideation without a specific plan, or a
suicide attempt or a specific plan for committing suicide.

Emotional distress. A constellation of uncomfortable
feelings that may include but are not limited to emotions such as

sadness, anxiety, tension, irritability, anger, hopelessness, guilt, and worthlessness.

Perception. "The elaboration, interpretation, and meaning given to a sensory experience . . . sensation [is] the more primitive, data-based experience generated by activities of receptors, while perception is the more brain-based interpretation of that sensory input" (Zimbardo, 1988, p. 144).

Physiological symptoms. A range of bodily discomforts which may include but are not limited to such symptoms as heart pounding, headaches, sleep and appetite disturbances, bodily aches and pains, and changes in sensory perception.

Reliability. "The consistency or reproducibility of test scores" (Crocker & Algina, 1986, p. 105).

Scale. Measurement instruments which are collections of items intended to reveal levels of theoretical variables not readily observable by direct means (DeVellis, 1991).

Self-rating scale. A procedure for data collection which allows the respondent to report information by placing a response or responses (beliefs/attitudes/feelings) to an anchor point or points on a survey scale.

Sensation. [The noun of sensory] According to Zimbardo, "the process of stimulation of a receptor that gives rise to neural impulses which result in an 'unelaborated,' elementary experience of feeling or awareness of conditions outside or within the body" (1988, pp. 143-144). Although there are eight senses--sight,

hearing, skin sensations, smell, taste, body movement, equilibrium, and organic sensitivity--for the purpose of this study, only the senses involving sight and taste will be of interest to this investigator.

Validity. "The extent to which the instrument adequately measures the concepts under study" (Green & Lewis, 1986, p. 101).

Summary

In this chapter, the background of the problem was explored. It was found that depression is costly to the health care system, causes significant morbidity and mortality, and tends to be unrecognized in primary care settings. The need for new valid and reliable depression scales which are practical to use in screening programs and which reflect current diagnostic standards was reviewed. Based upon the problems and the need, research questions and hypotheses for both the pilot study and the main study were generated, limitations of the study were cited, and research terms were defined.

The next chapter will be a review of the literature related to the study's variables as well as methodological concerns and issues in scale development and validation.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The purpose of this study was to develop a scale to measure
depression. Since an understanding of current research of the
construct under investigation was necessary, a comprehensive review
of literature pertaining to depression was appropriate. Therefore,
the following topic areas were reviewed and are presented in this
chapter: epidemiology and cause of depression, the nature and
clinical descriptions of depression, depression and the stress
framework, assessment of depression, and implications of the study
for health education. In addition, the review expanded upon
special issues in the use of the visual analogue scale which was
utilized in development of the Correa-Barrick Depression Scale.

Epidemiology of Depression

One important, often cited work is the National Institute of
Mental Health's Epidemiological Catchment Area (ECA) study by Eaton
et al. (cited in Goodwin & Jamison, 1990). This was a population
survey of five catchment area centers in the United States (New
Haven, Connecticut; Baltimore, Maryland; St. Louis, Missouri;

19

Piedmont County, North Carolina, and Los Angeles, California). A probability sample of over 18,000 adults (aged 18 years or older) living in the community was used. The communities studied varied considerably in size, population and geographic location. Although not completely representative of the American population, it was the most comprehensive epidemiological study to date (Goodwin & Jamison, 1990). Based on the ECA study, the risk factors for depression were age, gender, marital status, social economic status, and family history (Feighner & Boyer, 1991). The ECA study used a structured diagnostic instrument and the criteria for major depression as outlined in the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, revised (1987).

Incidence rate is the number of new cases of a disease in a population over a period of time, while prevalence rate measures the number of people in a population who, at a given time, have a disease or condition (Mausner & Bahn, 1985). The lifetime prevalence rate per hundred of major depression in the ECA study was 4.4 (Weissman et al., 1988) and 18.0 (Weissman & Mayers, cited in Feighner & Boyer, 1991). Prevalence was ascertained by asking respondents whether they had ever experienced any symptom during the time frame being sampled. Controlling for age distribution at each site, the investigators found there were no significant sex differences in the age of onset, which was 27 years.

One expectation about depression and age has been that the risk of depression increased with age; however, epidemiological studies have not supported this. The ECA results unexpectedly showed that major depression decreased with age; this finding was consistent across cultures (Feighner & Boyer, 1991). The decreased risk of depression with age cannot be fully attributed to artifacts such as reporting, recall, or mortality (Klerman & Weissman, 1989).

An increase was reported in the rate of depression among cohorts (that is, individuals grouped by some shared, continued temporal experiences [for example, the baby boomers who were born after World War II]). This association was consistent across cultures. The explanation for the unexpected findings concerning age was not very clear. One reason, however, might be the "birth cohort effect" discussed by Feighner and Boyer (1991) which was found to be very powerfully associated with depression for persons born after 1937 and was increased ten-fold for males born between 1957 and 1972. These birth cohort findings were also consistent with other temporal trends associated with depression, such as substance abuse and suicide, which tended to increase for young males born since the mid-1950s. Adolescents and young adults were reported to be increasingly depressed and to have experienced increased rates of alcoholism, drug abuse, suicide attempts, and deaths (Klerman & Weissman, 1989).

It was reported that women were twice as likely to become depressed as men; this association was consistent across cultures.

Married and never divorced persons had the lowest risk and those divorced or cohabiting had the highest rates. Persons with a family history of depression among first-degree relatives (that is, parents and/or siblings) had two to three times increased risk for depression (Feighner & Boyer, 1991).

In brief, family history, birth after World War II and especially between 1960 and 1975, and female gender were strongly associated with depression; these findings were replicated internationally.

One factor not associated with depression was race/ethnicity. According to Feighner and Boyer (1991), the similarities rather than differences among racial groups in rates of major depression in the ECA were striking. One exception was that a study in Taiwan showed a lower rate of major depression than found in Western countries. The differences may be truly racial or may be due to cultural differences such as the selective underreporting of some types of symptoms (Feighner & Boyer, 1991).

The ECA study also found no association between social economic status and major depression; however, rates of major depression were lower among employed and financially independent persons. The unemployed and those on public assistance had a threefold increased risk of major depression. Lack of employment posed psychological stress that could contribute to developing or maintaining major depression (Feighner & Boyer, 1991).

According to another epidemiological study by Robbins et al. (cited in Goodwin and Jamison, 1990), there were no significant differences in educational achievement and depression across three different areas--urban, suburban and rural. There were, however, significant but conflicting differences in urban versus rural rates of major depression. Because the epidemiological evidence for urban or rural risk was even, no conclusion was drawn at that time.

Causes of Depression

From research studies that examined genetic, biological, and environmental factors associated with depression, many theories about its cause or causes had been developed. Family history was strongly associated with depression, and this finding spawned genetic studies. However, no one cause has been elucidated (DePaulo & Ablow, 1989; Mondimore, 1990; Goodwin & Jamison, 1990; Feighner & Boyer, 1991).

The view that there existed a physiological basis for depression was first conceived almost thirty years ago by Schildkraut (1965) who reviewed evidence that formulated the catecholamine hypothesis of affective disorders. The physiological basis for depression will be reviewed since it sets a biological framework for subsequent discussions on physiological concepts of stress and depressive illness.

Animal models showed that iproniazid, a monamine oxidase inhibitor (an antidepressant), produced behavioral excitation and

higher brain levels of norepinephrine by preventing one of the catecholamines from breaking down serotonin in animal species. Pharmacological studies with reserpine produced sedation in experimental animals; this was thought to result in depletion of norepinephrine stores, dopamine, and serotonin resulting in depression (Schildkraut, 1965). The research methodology utilized involved administration of reserpine over a period of time and the measurement of catecholamine levels. A significant correlation was found between the behavioral effect of sedation in animals and depletion of the catecholamines. Moreover, there was a temporal correlation between the return of normal behavior, as assessed by motor activity in these animals and restoration of norepinephrine and serotonin. Studies using amino acid precursors of catecholamines and serotonin administered to animals that had been given reserpine also replicated the previous experiments. Specifically, there was a return to normal behavior after administration of a catecholamine precursor. Pharmacological effects of reserpine on man have produced depression, which would also seem to support the hypothesis. Other support for the hypothesis evolved from studies using animal models in which imipramine, an antidepressant, prevented or reversed reserpine-like sedation in animals. Imipramine worked by limiting the access of norepinephrine to mitochondrial monamine oxidase. These studies suggested that too little norepinephrine led to depression in humans, while too much led to mania.

One of the issues raised by Schildkraut (1965) was whether the antidepressant drugs corrected or cured the abnormal state in depression or whether the drugs compensated symptomatically for the abnormal state. Proof of the hypothesis depended upon direct demonstration of the biochemical abnormality in the naturally occurring illness. Since true experimental laboratory research was methodologically not feasible in human subjects, the question remained unanswered as to whether animal research could be extrapolated to humans.

Other research had been done to examine biological correlates of affective disorders. Studies of depressed patients were done to measure urinary concentration of the catecholamines, the hypothesis being that lower values would be evident during depressed episodes in human subjects. Lower levels of urinary norepinephrine were found in depressed patients compared to controls (Depue & Evans, 1981). This hypothesis was methodologically flawed since urinalysis of catecholamines can be influenced by physical activity of patients. Unfortunately, urinary studies were not direct measures of catecholamine metabolism in the central nervous system.

Frazar, Hancock, Mendels, and MacIntire (cited in DePue & Evans, 1981) hypothesized that depression was not a result of the amount of catecholamines, specifically norepinephrine in the brain, as Schildkraut thought, but rather a problem at the receptor site. For this reason, cyclic adenosine monophosphate (AMP) has been studied in terms of its relationship to depressive disorders (Depue

& Evans, 1981). Most hormone and biogenic amines combine with a specific receptor at the membrane of the target cell. This binding causes activations of the enzyme adenylate cyclase, a protein, which then catalyzes the transformation cell to cyclic AMP which then initiates any number of cellular functions (Berne & Levey, 1983). Different values in the urine were found in depressive versus normal patients, but changes in physical activity also tended to show changes in urinary AMP.

Much of the biological research on depression has focused on biogenic amines, specifically the catecholamine, norepinephrine. This focus actually represented a small number of neurotransmitters since there are many different nervous system transmitters. Research had been conducted on neurotransmitters but no conclusions were reached (Teuting & Koslow, 1983). Other variables studied included electrolytes, REM sleep patterns, electroencephalogram, biorhythm, melatonin, and neuroendocrine functioning.

During the 1980s, much research interest centered around the neuroendocrine system. This was due to several reasons: (a) the relation of many of depression's symptoms to hypothalamic influence (functioning of the hypothalamus is dependent upon the biogenic amines), (b) the higher incidence and prevalence of depression among women, (c) mood changes in women, and (d) consistent, significant elevations of cortisol in depression which has been replicated in other research studies (Depue & Evans, 1981; Teuting & Koslow, 1983).

Endocrine disorders may be produced by changes in hypothalamic activity which can also be altered by stress (Berne & Levey, 1983). To make the issue more complex, hormones affect the amine uptake and metabolism, the electrical excitability of nervous tissue, the distribution of sodium and potassium, and the sensitivity of receptors. To study the role of neurotransmitters in depression, one is compelled to study the role of the endocrine system as well, especially the hypothalamic, pituitary, adrenal axis (the HPA).

Another interesting aspect of depression and mania is its occurrence in patients with Huntington's chorea (a rare, genetic disease that produces deterioration in cognitive function), stroke, and Parkinson's disease. According to DePaulo and Alblow, researchers ". . . have found that injuries to the brain might cause depression by destroying nerves that use the catecholamine transmitters. . . ." (1989, p. 25). Endocrine abnormalities have been linked with affective disorders. Diseases such as hypothyroidism produced depression, while Cushing's disease, which results from excessive stimulation of the adrenal glands by the pituitary, results in an overproduction of cortisol. Some patients suffering from Cushing's disease have also developed symptoms of manic-depression.

One of the most interesting studies to date addressed the role of genetics in affective disorders, as several researchers have discovered genetic patterns. The Amish study was cited as a

landmark in locating a genetic marker on chromosome 11 in the Amish population (Egeland, 1987). This group was ideal to study because pedigrees were readily established with a minimum of error in the Amish families. Unfortunately, Egeland's findings could not be replicated.

In the 1990s, depression research has begun to be directed toward discovering how genes alter brain chemistry. Genetics, however, does not seem to explain all of it. In every genetic study, the concordance rate for monozygotic twins who developed affective disorders ranged from 50% to 90% (DePaulo & Ablow, 1989). What this meant was that, if one identical twin developed manic-depression, the other twin had a greater than average chance of also developing the illness. If the concordance rate fell below 100%, then the other twin had less than a 100% chance of developing the disease. If the disease were of complete genetic causation, a 100% concordance rate would be expected.

Other factors such as environment have clearly played a role in triggering the disease in those who were genetically susceptible, and much has been written about the psychological theories of depression. Therefore, it was prudent to review all of the major competing theories. The learned helplessness theory viewed depressed individuals as having experienced a loss of control over their environment. While this may be true, it does not explain the cyclical nature of depression or manic-depression. The cognitive theorists believed that negative thinking led to

depression, but this could also be a symptom of depression rather than the cause (DePaulo & Ablow, 1989).

Although not fully developed and tested, current models on manic-depression tended to view it in the disease paradigm. The practical experimental constraints of studying the human brain have prohibited the kind of direct experimental research required to flush out the biological pathways.

Stress and Depression

For the purpose of this discussion stress was defined as ". . . a physiological reaction, or response, regardless of the source of the reaction. . . . stressor will be used to refer to the stimulus that serves to engender the stress response" (Everly, 1990, p. 6). According to Everly, stressors may be psychosocial or biological.

There was little agreement among researchers about the conceptual and operational definitions of stress. Most studies reviewed examined stress in terms of life events. Stressful life events were studied by O'Connell (1986) who found a relationship between such events and the onset of depression and bipolar disorder. Research on linking stressful life events and the onset of illness were found to have correlation coefficients less than .30. Previous stress in childhood, specifically childhood loss of a parent, was not substantiated by research (Golderber & Breznitz, 1982). Although it is known that stress can cause disease by

altering many systems in the human body including neurotransmitters and endocrine functions of the brain (Allen, 1983), studies by Ursano, Boydstun and Wheatley (cited in Goldberger & Breznitz, 1982) on returning Vietnam prisoners-of-war did not show a higher incidence of psychiatric disorders, nor did Clayton (cited in Goldberger & Breznitz, 1977) find any significant increase in clinical depression after bereavement.

Glassner and Haldipur (1983), who studied 46 subjects with bipolar disorders found an association between life events, as measured by the Holmes and Rahe scale, and onset of the illness. Likewise, Bidzinska (1984) conducted a controlled study of 97 patients with affective disorders and 100 healthy control subjects, and the findings were consistent with those of Glassner and Haldipur (1983). Bidzinska found that acute and chronic stress factors occurred more in the group of patients with affective disorders than among the control group over a similar period of time. Stress factors were investigated with the Life Events Questionnaire (Bidzinska, 1984) utilizing a retrospective design. Types of events that were associated with affective illness included marital/family conflicts, health problems, emotional and ambitional failures, lack of success, and work overload.

Stressful life events were also studied by O'Connell (1986) who developed a model on the relationship between such events and the onset of depression. O'Connell's model described the

interaction between genetics and stress and development of the illness based on one's psychobiological vulnerability.

Ambelas and George (1988) studied the concept of the meaning of life events for patients and observed that meanings were quite specific for each patient. Dunner, Patrick and Feive (1979) found that, between the initial or subsequent episode of affective illness in a sample of 79 bipolar patients, about half of the patients recalled a life event in the three month interval before their initial episode. Their data suggested that life events were associated with illness onset. The study's limitation, however, was the case history design.

Paykel (1986) eloquently summed up the research dilemma:

While evidence for a genetic element is strong, that for detailed neurochemical abnormalities, even in terms of biogenic amines is weak, and the prospect of a specific molecular genetic abnormality or set of abnormalities may still be some way off. Failing these, all attempts to explain the crucial cycling element that characterizes the disorder still seem speculative. (p. 264)

Clayton (1986) concurred with Paykel and summarized, "Reading the literature, one is struck by the fact that there is no single finding that invariably determines or predicts an association" (p. 265).

Two recent studies examined stress as a physiological measure and its association with affective illness. Roy, Guthrie, Pickar and Linnoila (1987) experimentally studied depressed patients and controls for plasma norepinephrine responses to cold challenge. Subjects had their hand placed in ice cold water for one minute,

while the researchers studied the effects of a physical stressor (ice water) on the blood level of norepinephrine. They found that depressed patients showed significantly higher plasma norepinephrine levels than did the control subjects and concluded that there was a dysregulation of the noradrenergic system in depression. This study suggested that the effects of stress, measured at the physiological arousal state, might be one explanation for the genesis of depression.

Another study found a relationship between thyroxine levels and recovery rates in depression (Southwick, Giller & Kosten, 1989). Specifically, it was found that patients who were least likely to improve clinically were those with initial thyroxine levels in the low normal range and whose levels either stayed the same or decreased.

One Russian study (Kamenskaya & Mikhailova, 1982) found that, under conditions of stress delivered as an expectation of electric shock, a group of affective patients had pulse rates decrease in relation to baseline compared to controls who had an increase of 60% to 70% in relation to baseline. It was concluded that parasympathetic responses clearly dominated in depressive patients and that the posterior hypothalamus was involved in the pathophysiological mechanism in depression.

Research methodology problems have plagued most of the life stress studies. Specifically, weaknesses were apparent on the reliability of retrospective reporting, self-reports, and the

temporal association of the event to the onset of illness. Conceivably, the depression itself could create life events such as marital problems, and job loss. Given the present state of the research, there were too many puzzles. One of the most difficult methodological problems was discerning whether one was studying the cause or effects of affective illness. Another methodological problem centered on the different ways of classifying and measuring depression, particularly prior to the development of diagnostic criteria.

Depressive illness was found to be conceptually compatible with a stress framework. Research data suggested there was inadequate evidence to conclude that stress alone was sufficient or specific to causing depression. Stress seemed to be a necessary but insufficient component in triggering depression. When stress-depression associations were found, the connections were either inconsistent or the correlation coefficients were not impressive. The contradictory findings posed a puzzling picture in understanding etiologic phenomena. In conclusion, the best model to explain the onset of a major depression was that stress was a risk factor for the precipitation of a major depression in biologically predisposed individuals.

The Experience of Depression

The experience of depression ranges from feeling blue to feeling that life is so unbearable that suicide is attempted.

Clearly, depression is a global construct that causes pervasive changes of varying degree and severity. Descriptions of the criteria for depression were provided by the revised Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R) of the American Psychiatric Association (1987), and they have remained important elaborations of some of the phenomenological aspects of the depression experience. The DSM-III-R reported that "a person with depressed mood will usually describe feeling depressed, sad, hopeless, discouraged, 'down in the dumps,' or some other colloquial equivalent" (p. 219). The depressed person may deny depression, but it can be inferred from observations about the person looking sad or depressed. The DSM-III-R further elaborated upon psychomotor manifestations:

> Psychomotor agitation takes the form of inability to sit still, pacing, hand-wringing, pulling or rubbing of hair, skin, clothing, or other objects. Psychomotor retardation may take the form of slowed speech, increased pauses before answering, soft or monotonous speech, slowed body movements, a markedly decreased amount of speech (poverty of speech), or muteness. A decrease in energy level is almost invariably present. . . . The smallest task may seem difficult or impossible to accomplish. (p. 219)

Sleep involves excessive sleeping which may take the form of daytime sleepiness or taking excessive naps. At times, depressed persons will initially seek help for the sleep disturbance rather than the depression.

The experience of depression can vary in cultures. In Eastern and African societies, many more depressed persons reported bodily symptoms; whereas, depressed persons in Western cultures

reported more dysphoria and guilt. It was also reported that men presented fewer symptoms of depression than did women, even when the level of social or occupational impairment was the same (Angst & Dobler-Mikola, 1984).

Depression may often be masked or hidden in despondent persons. Masked depression included those manifestations of a depression in which the primary symptoms were physical. The most common symptoms found by Lopez-Ibor (1972) were (a) aches and pains, (b) psychosensorial disorders (dizziness and vertigo, visual disturbances, and disturbances in the perception of space), (c) neurological symptoms (inability to remain quiet or a need to be constantly moving) and restless legs syndrome (leg jitters/fidgets), and (d) psychosomatic symptoms (gastrointestinal, respiratory, genitourinary, cardiovascular, skin, obesity, and thinness).

Clinical Descriptions of Depression

Descriptions of depression by psychiatrists provided the investigator with behavioral manifestations of depression. Anecdotal descriptions by persons with depression provided the investigator with rich, qualitative aspects of depression that an inventory of symptoms could not provide. Both sources of anecdotal descriptions thus provided the investigator with a balanced perspective for developing a depression scale.

Clinical descriptions of depression were well identified by

Goodwin and Jamison (1990):

> Mood, in all of the depressive states, usually is bleak,
> pessimistic, and despairing. A deep sense of futility is
> often accompanied, if not preceded, by the belief that the
> ability to experience pleasure is permanently gone. The
> physical and mental world are described as monochromatic, as
> shades of grays and blacks. Heightened irritability, anger,
> paranoia, emotional turbulence and anxiety are common
> correlates of depressive mood. (p. 36)

Information about the relationship of sensory/perceptual

phenomena and depression was limited to a few textbook references

and descriptions of patient experiences. A comprehensive review of

research studies concerning the clinical descriptions and symptoms

of depression failed to retrieve research data about sensory

alterations during depression, although it was reported that there

is heightened perceptual awareness in mania (Goodwin & Jamison,

1990). Nonetheless, there have been qualitative descriptions on

sensory awareness disturbances. Campbell (cited in Goodwin &

Jamison, 1990) described the depressed person's decreased sensory

awareness as "In addition to distortions in sensing impressions

such as a queer, odd or unreal feeling, the patient may complain of

a universal dulling of the emotional tone" (p. 36). Kraeplein's

work (cited in Goodwin & Jamison, 1990) provided a description of

the disturbed sensory aspects of depression on music awareness:

"Everything has become disagreeable to him; everything wearies him,

company, music, travel, his professional work" (p. 37). One

patient vividly described sensory experiences: "Everything I see,

say, or do seems extraordinarily flat and pointless; there is no color, there is no point to anything" (Goodwin & Jamison, 1990, p. 41).

According to Goodwin and Jamison (1990), in depression, cognition was found to be markedly slowed; suicidal thinking was often a dangerous potential, and rumination and hypochondriacal thinking were present. The cognitive experience was described as follows:

> Thinking is difficult to the patient. . . . He cannot collect his thoughts or pull himself together; his thoughts are as if paralyzed, they are immobile; he has no longer command of knowledge formerly familiar to him, he must consider a long time about simple things, he calculates wrongly, makes contradictory statement, does not find words, cannot construct sentences correctly. (Kraeplin, cited in Goodwin & Jamison, 1990, p. 38)

Activity and behavior were also reported to be disturbed in depression and may be slowed in such a way that the behavior was observable: "The depressed individual usually walks slowly and reacts sluggishly. He appears to push himself along, as if he were being held back, rather than propelling himself with normal agility" (Campbell, cited in Goodwin & Jamison, 1990, p. 38).

Sleep was reported to be disturbed in a depression. F. Scott Fitzgerald (cited in Goodwin & Jamison, 1990) described his experience ". . . hating the night when I couldn't sleep and hating the day because it went towards night" (p. 41).

Suicidal depression can be a serious consequence of depression, and one of its most poetic descriptions was given by Alvarez (1973):

A suicidal depression is a kind of spiritual winter, frozen, sterile, unmoving. The richer, softer and more delectable nature becomes, the deeper that internal winter seems, and the wider and more intolerable the abyss which separates the inner world from the outer. (p. 79)

Assessment of Depression

There were reported two primary methods for assessing depression: the interview/observer method and the self-report. The interview/observe method was not a panacea for assessing depression, as there were problems associated with its use. In addition, it required proficiency by the interviewer, usually a psychiatrist. Since the clinician was the source of data, the data was subject to the clinician's judgment before it was reported. Additionally, there may be differences in the interview results due to the training of the rater, the theoretical base of the rater, and the role of the rater (Lambert, Christensen, & DeJulio, 1983).

Depressive disorders in the population were usually assessed by one of the following: structured diagnostic interviews, clinician rating, and self-ratings.

Structured Diagnostic Interviews

Structured diagnostic interviews were standardized interviews that specified the questions and inquiries made by raters and the response options for patients. It also required interviewer selection and training. There were four major standardized diagnostic interviews (Hasin & Skodol, 1989): (a) the Structured Clinical Interview for DSM-III-R; (b) the Schedule for Affective Disorders and Schizophrenia (Endicott, Spitzer, Fleiss & Cohen, 1976); (c) the Diagnostic Interview Schedule (Robins, Helzer, Croughan, & Ratcliff, 1981), and (d) the Present State Examination (Wing, 1970).

Clinician Rating Scales

Clinician rating scales were similar to diagnostic interviews but differed in that they were less structured. According to Rabkin and Klein, clinician rating scales ". . . require the interviewer to cover specific areas with the client and they provide more or less specific response options, but the questions addressed to the patient are not specified" (1987, p. 48). They are subject to variance between raters. Coyle (1990) reported that interrater reliability could be increased when definitions were given and anchor points provided on the scale. One of the most popular clinician rating scales found was the Hamilton Rating Scale for Depression (HRSD), a clinician rating scale requiring trained personnel (Folstein & Lurian, 1973). The HRSD has remained the

standard to which all other rating scales were compared (Rabkin & Klein, 1989) and has been cited extensively over the past several decades in research studies of depressive disorders (Cronholm & Daly, 1983).

The HRSD quantified severity of illness based upon information given by the patient to a trained clinician during an unstructured interview. A 21-item rating scale, its interrater reliability was reported as ranging from .80 to .90 in the original psychometric work completed by Hamilton (1960). Since that time, Hedlund and Vieweg (1979) reported interrater reliability coefficients above .85 in seven of eight studies they reviewed and internal consistency reliability estimates from .83 to .94.

Self-Report Measures

There were two general types of self-report measures: (a) inventories (Mayer, 1978), which consist of a series of graded items in which the subject rates the presence, frequency, or intensity of a range of symptom complaints and (b) checklists (Coyle, 1990) which were a series of ungraded items.

The self-report procedure was developed by Woodworth (cited in Derogatis, Lipman, Uhlenhath & Covi, 1974) from the need to process large numbers of men for military service during World War I when a psychological scale, the Personal Data Sheet, was developed as a self-report measure. There were advantages to using self-rating in assessing depression, as it provided a simple method

of assessing a person's condition and could show changes in the severity of their depression (Hamilton, 1967). Self-rating (a) was effective in initial screening and removed observer bias (Deforge & Sobal, 1988); (b) had highly significant correlations with observer ratings (Carroll, Feinberg, Smouse, Rawson, & Greden, 1981); (c) offered a valuable means of detecting depression as limited resources restricted the use of the interview method which required more time and effort (Deforge & Sobal, 1988), and (d) permitted an eclectic interpretation among interdisciplinary staff (Lambert, Christensen & DeJulio, 1983).

Self-ratings did not require skilled interviewer technique and were potentially easy to administer, score, and interpret. Self-reports were not prone to problems in reliability as were structured interviews and clinician ratings, since key issues on self-reports indicated that it required internal consistency and stability over time. Another major advantage was their ability to quantify subjective symptoms for subsequent analysis (Coyle, 1990).

There were limitations with the use of self-report measures: they were found to be prone to subject faking (Anastasi, 1988) and to deliberate misrepresentation and exaggeration, extremeness, and other response sets (Rabkin & Klein, 1989). Gove, McCorkel, Fain, and Huges (1976) reported three response sets with self-reports: nay saying, perceived trait desirability, and need for approval. However, Gove et al. explained that these response sets did not

invalidate the measures since error was not due to systematic error but due to random noise.

Self-rating scales for assessing depression have diverse health education application, and these will be discussed later in this section. However, applications in mental health settings included initial and follow-up measures of the signs and symptoms of depression in clinical drug trials and in evaluating patient responses to therapy, as well as an adjunct to the clinical interview.

In primary care settings, self-reports were found to be valuable in helping the busy primary care physician, with limited time and resources, to identify and diagnose depressive disorders; thus, the self-rating scale became a "depression thermometer" (Zung, 1990).

Conversely, disadvantages to the self-rating scale also existed. Denial, exaggeration, and loss of insight are unavoidable during a depression (Carroll, Feinberg, Smouse, Rawson, & Greden, 1981). Since self-rating scales are high in sensitivity but low in specificity, they may pick up a false positive because they overestimate depression due to emphasis on symptomatology rather than diagnosis (DeForge & Sobal, 1988). Other problems reported with use of self-ratings were masking and accuracy issues associated with a patient's awareness, literacy, and motivation (Lambert, Christensen, & DeJulio, 1983). Self-reports were also plagued by problems with social desirability (that is, when an

individual was motivated to present himself or herself in a positive way) which distorted the item responses (DeVellis, 1991). Lastly, a person with severe depression may not even be able to complete the scale.

One canon about self-report is that it should never be used for diagnosing depression without the use of multiple measures such as (a) a physician's physical examination to assess the presence of other diseases, (b) determination of the duration of depression, specifically whether longer than two weeks, and (c) professional assessment.

Measurement of Trait versus State

It became necessary to discuss the concept of trait versus state since both were relevant to this study and because depression can be conceptualized as either a state or a trait, although each concept differs.

Human traits are human tendencies to behave in a certain way (Campbell, 1963). States exist at a given moment in time and at a particular level of intensity. They can last when a stimulus persists, but they are usually transient. If a state is relatively transitory, then a trait is enduring (Spielberger, 1972). Traits and states may change, but states change more often. There were certain caveats for measuring states and traits. When measuring states, the instructions should ask respondents to describe how they feel "today." In contrast, when measuring trait, the

instructions should ask respondents how they usually feel. This distinction was important for measuring depression as a perceived state.

Self-Report Depression Scales

Depression self-report scales with known validity and reliability are available, and each scale will be reviewed historically, starting from the earliest date and proceeding in chronological order to the most recent. Appendix C summarizes the main psychometric results.

Visual Analogue Scale

The Visual Analogue Scale (VAS) began to be utilized in 1921 when Hayes and Patterson used it as a graphic rating scale for measuring feelings in their psychiatric practice (Bond & Lader, 1974). Freyd (1923) cited several advantages to the VAS: (a) it was easy to complete, (b) it did not require much subject motivation, and (c) the researcher could make however fine a resolution on the scale as desired. Joyce (1968) believed that the VAS reduced the difficulties of response sets.

The VAS was not used again until Aitken (1969) promoted its application as the shortest self-report depression measure. Subjects rated their depression by selecting a point on a continuous line representative of depressed mood. Anchor end points were often opposite such as not at all depressed to very depressed. According to Aitken (1969), "The paucity of suitable

quantitative terms in common speech limits the amount of information which can be transferred. . . . For the measurement of feelings, communication based on a simple visual analogue seems appropriate" (p. 989). Aitken reported that people tended to like the VAS since it was not stressful to respondents who found their feelings between two discrete categories. The main argument for using the VAS was not how it could discriminate feelings down to the millimeter, but that it freed the researcher and respondent from being "boxed in" by specific categories. Although traditionally used as a continuous line without gradations, the VAS has recently begun to be used with them.

Bech, Kastrup and Rafaelson (1986) noted one disadvantage to the VAS: it assumed interval scaling of responses even though numerical scaling was not formatted on the scale. This investigator believes that such a problem can be solved, since interval scaling can be facilitated by using numbers on the line with specific directions to the respondent to circle the appropriate number on the continuum.

Folstein and Luria (1973) reported that reliability and validity of the Visual Analogue Mood Scale (VAMS) was demonstrated in both a military and a private psychiatric hospital using inpatient populations. A VAMS was constructed on a rectangular card (100 mm by 35 mm) on which the following was printed: "How is your mood right now? A mark on the line toward the left represents your worst mood, toward the right, your best." The VAMS score was

determined by measuring the distance in millimeters from the left end of the card to the patient's mark. The VAMS correlated with concurrent validation measuring both the affective and nonaffective patient groups using the Self-rating Depression Scale (SDS) and the Clyde Mood Scale.

Selth (1990) investigated the assessment of clinical depression in a geriatric nursing home population using a behaviorally anchored rating scale, the Behavioral Observation of Depression in the Elderly Scales (BODES), and a one-item, self-report visual analogue scale of depression (VASD) which consisted of a "continuous 6.5" horizontal line representing the continuum of depressed affect, with endpoint labels of 'not depressed' and 'very depressed,' and evenly spaced numerals 1 through 7 placed below the line" (p. 129). Nursing home residents were asked to select a point somewhere along the continuum after the words of the VASD were read aloud to them. Results suggested good convergent validity with the VASD and other depression scales (the self-report Geriatric Depression Scale and the interviewer-rated Montgomery-Asberg Depression Scale). Selth determined a moderately high and statistically significant positive correlation when comparing the VASD to the GDS (.68) and to the MADRAS (.69). The VASD accurately discriminated between subjects with antidepressant medication versus those with no medication.

Selth further reported favorable psychometric findings on the VAS and initially hypothesized that the VAS would not be effective

in measuring depression in a geriatric population since they often minimize self-reports of depressed mood. Unexpectedly, resident subjects selected a variety of VASD scores. The study's positive findings regarding the VASD seems promising given its brevity and its familiar and simplistic format. Selth concluded that the VASD could serve as a useful, brief, and easily administered alternative self-report scale of depression for elderly clients.

Issues with the Visual Analogue Scale

Gift (1989) addressed several disadvantages with the VAS. For example, the use of gradations on the VAS reduced its sensitivity. In addition, since the VAS only quantified intensity, researchers did not always agree on the wording of the anchors to reflect the extremes of a feeling or sensation. For instance, depression anchors could vary from not at all depressed to most depressed I have ever been versus so depressed that I can't stand it versus depressed enough to want to kill myself. Another disadvantage was response set which occurred when multiple horizontal scales were used; that is, subjects tended to mark all the scales down the middle. Thus, raw data from the VAS may be skewed, but one way noted to correct this was with the arcsin transformation to normalize the distribution of scores (Aitken, 1969). However, according to Gift, comparisons between the results obtained with and without transformed scores showed no differences in sensitivity or results obtained. Gift also pointed out a major advantage of

the VAS: vocabulary level of subjects did not have to be of major concern since the scale was visually oriented rather than language oriented. However, there is one final word of caution when using the visual analogue scale in elderly populations: Carlson (1983) found that the VAS may be a troublesome measure of change over time due to problems with memory recall, especially for some elderly who may have difficulty remembering prior experiences.

When the VAS was used for repeated measures, the issue of whether subjects should see their previous responses was debateable. Guyatt, Berman, Townsend, and Taylor (1985) concluded that letting patients see their previous responses would result in reproducibility of the scale. Scott and Huskisson (1979) assessed pain severity in patients with rheumatic disorders and compared pain measurements made by patients with and without prior knowledge of responses on the VAS. Since patients tended to overestimate their pain severity when previous scores were not available, it was recommended that prior scores should be made available when serial measurements of pain were made on long-term experiments. Joyce, Zutshi, Hrubes, and Mason (1975) found that the VAS was more satisfactory than a four-point scale for patient self-rating of pain intensity; moreover, patients preferred the VAS.

One of the most compelling reasons to use the VAS was its purported sensitivity to change. As such, the VAS was shown to be useful in detecting drug effects in normal subjects (Bond & Lader, 1974). The issue of whether there was evidence to support that the

VAS was superior to other scales remained debatable, as research results were mixed. Davies, Burrows and Poynton (1975) compared overall scores on the Beck, Hamilton, Zung, and VAS for a group of depressed patients. They found highly significant correlations of all the depression rating scores at days 0, 7, 14, and 21 and reported that the VAS was as useful in quantifying symptoms as were other instruments.

Guyatt, Townsend, Berman and Keller (1987) compared a seven-point Likert with a VAS in a questionnaire measuring quality of life in chronic lung disease. It was found that the two methods showed comparable results and recommended the Likert scale because there was less instruction time involved in teaching patients how to complete it.

Sriwatanakul et al. (1983) found that the horizontal scale with gradations was most preferred by volunteer subjects. Results of the study suggested that visual analog scales may be more sensitive than descriptive pain scales in a postoperative patient sample, since significant changes of pain intensity on the VAS were seen in the absence of changes in the verbal scores, and these changes occurred at a time when the effects of the pain medication were at peak. Joyce et al. (1975) also found the VAS to be a more sensitive measure of subjective sensation than a four-point rating scale.

The VAS was recommended in situations when sensitivity was desirable or essential (Gift, 1989). Sight impaired individuals

could use it when the anchors were large enough or were read aloud. One major disadvantage to the VAS was that people found it difficult to transcribe a subjective feeling to a straight line. This could be overcome, however, by producing written instructions at the top of the scale. Gift reported that measurement on retest could be a problem to some subjects who may have problems with recall.

The VAS has been used in a variety of research settings including areas that concerned pain, sleep, and other feelings and sensations beyond mood. Aitken (1967, cited in Aitken, 1969) used the VAS to measure the degree of apprehension felt by fighter pilots who were presented with 100 mm lines with the extremes defined as maximal relaxation and maximal panic. In a clinical trial of two hypnotic drugs and a placebo, the quality of sleep was assessed using a VAS (Aitken, Southwell, Wilmhurst, cited in Aitken, 1969).

There have been several studies to validate the VAS with depressed persons. In use with depression, the VAS has usually been composed of a single line, about 10 centimeter long, with the wording as depressed as I have ever been to not at all depressed. Reliability and validity with this scale has been poor, with ranges from a high of .78 with the Hamilton Depression Rating Scale (HDRS) to a low of .06 with the HDRS (Thompson, 1989). In another study, Cella and Perry (1986) studied 34 family members of patients who had been admitted to a large urban burn center. The relatives were

asked to complete three 100 mm visual analogue scales. Subjects were instructed to indicate with a mark on the line how they felt right now using three sets of polar statements placed on the 100 mm line: not at all depressed to most depressed I've ever felt, not at all nervous to most nervous I've ever felt, and not at all stressed to most stressed I've ever felt. Three scales were used to establish convergent validity with standardized measures for depression (the Beck Depression Inventory), for anxiety (the Spielberger State-Anxiety Inventory), and for distress (the Perceived Stress Scale). Pearson correlations among the three visual analogue scales showed significant association with the Profile of Mood States but not with Perceived Stress Scale. The researchers suggested that the VAS measured something closer to general distress than appraised stress. In repeated-measure analysis of variance for each analogue scale, significant changes over time were seen in depression and anxiety but not in distress. Validity was demonstrated by significant correlations between the depression analogue and the Beck Depression Inventory, the anxiety analogue and the Spielberger State-Anxiety scores, and the distress analogue and the Profile of Mood States total scores. Cella and Perry (1986) concluded, "although visual analogue scales appear to be simplistic and superficial, they seem capable of measuring feeling states in a quick, reliable, and relatively sensitive manner" (p. 831). Although their study supported the use of visual-analogue scales for rapid assessment of feeling states when

more lengthy scales were infeasible, Cella and Perry cautioned the user against using the single-item VAS to measures complex constructs on anxiety, depression, or general distress.

Use of Vertical or Horizontal Visual Analogue Scale

Dixon and Bird (1981) investigated reproducibility along a vertical 10 cm VAS and found there was a tendency for respondents to estimate too high on the scale. Eight normal volunteers were each presented with a series of 10 vertical, 10 cm reference lines seven times. Results indicated that the most troublesome positions to reproduce appeared to be in the region of the midpoint, with good reproducibility occurring near the apices and at the center. There was a tendency to estimate positions too high on a vertical VAS when access to the initial reference line was available and concluded that subjects should not see previous responses. Furthermore, Dixon and Bird found an additional source of error on the vertical VAS in that the angle at which the scale was viewed may cause visual distortions for the respondent. Scott and Huskisson (1979) reported that vertical and horizontal visual analogue scales showed high correlation scores ($r = .99$, $p < .001$), but scores from horizontal scales tended to be slighter lower than those from vertical scale. It was noted that it would be essential for scales to remain identical during any study.

Visual Analogue Scale Sensitivity

Several studies validated use of the Visual Analogue Scale
(VAS) for measuring changes in both mood and pain sensations in
patient populations, and these studies are summarized.

Zealley and Aitken (1969), in a study of 13 patients with
depressive illness, found good correlation coefficients with three
methods of depression assessment: (a) psychiatrist overall rating,
(b) the Hamilton rating score, and (c) the VAS. All correlations
ranged from .79 to .90 on admission and were significant. In a
study of clinical trials of antidepressant drugs using the VAS, 10
depressed outpatients were randomly assigned to either a new or an
established antidepressant drug on a double-blind basis to test
whether the new drug worked faster. Results showed that the
regression slope for the established drug was highly significant,
but there was no significant slope for the new drug, thereby
rejecting the claim that the new drug was superior. The
researchers claimed that, for the assessment of mood in depressed
patients, the VAS had been shown to be practical, reliable, and
valid and was suitable for the measurement of change. This study
would have been strengthened if comparison between the VAS and
other self-report measures had been used.

Little and McPhail (1973) studied measured mood at monthly
intervals in eight female outpatients diagnosed with long-standing,
recurrent depressive illness. The researchers used the following
procedure: At each patient's monthly visit, two psychiatrists

independently but simultaneously marked their own copies of the VAS, while the patient completed a VAS and Beck Depression Inventory. The study was conducted over a 16-month period. Results indicated correlations between the three measures were high and significant. In their discussion, the researchers reported that the VAS gave a reliable and valid measure of mood change over time. Little and McPhail also concluded that the Beck Depression Inventory was less sensitive on The profile showed outpatients with only mild or moderate depression.

Luria (1975) studied 62 patients with psychiatric disorders to validate the Visual Analogue Mood Scale (VAMS; Folstein & Luria, 1973). Two types of observational data were collected. The treating physician, blind to the mood slip scores, rated weekly each of his patient's overall clinical condition on a 100 mm analogue scale. Concurrent validity with the Clyde Mood Scale (Clyde, cited in Luria, 1975) and the Self-rating Depression Scale (Zung, 1965) was tested and significant correlations were found. In addition, the VAMS showed validity with observed behavior using a nurse's rating scale (the Psychotic Inpatient Profile) and with observed changes in overall clinical conditions using physician ratings on an analogue scale. Reliability of the VAMS was significant both across patients and within patient groups.

The VAS has also been validated in studies measuring intensity of pain. Price, McGrath, Rafii, and Buckingham (1983) provided direct evidence for ratio scaling properties of the VAS in

30 chronic pain patients and 20 healthy volunteers. The
researchers were interested in a ratio rather than an interval
score to enable comparisons between different types of pain and
interpretations of analgesic efficacy. Each subject participated
in two experimental sessions in which six intensities of contact
heat varying from 40 to 51 degrees centigrade were applied to the
ventral forearm. Patients made VAS responses to both the sensation
intensity and affective magnitude of both experimental heat pain
and chronic clinical pain. The researchers found that subjects who
experienced a heat pulse stimulus (heat delivered by a hand-held
contact thermode which was applied to the ventral forearm of
subjects for five seconds duration) perceived pain twice as intense
as the standard stimulus (43 degrees Centigrade). Findings
indicated that the observed values coincided with predicted values
on a linear regression line which supported the ratio scaling
properties of the VAS.

Joyce, Zutshi, Hrubes and Mason (1975) compared a VAS and a
four-point scale in patients suffering from chronic pain. The
purpose of the study was to assess the effect of the subjects'
access to previous ratings and to assess the efficacy of two doses
of two commonly used mild analgesics. For four weeks, using an
experimental design, each of the 74 patients in the study sample
was asked to record pain intensity following drug treatments. The
four-point scale was anchored using no pain at all, some pain,
considerable pain, and pain which could not be more severe. The

VAS was a 10 cm line with extremes marked <u>I have no pain at all</u> and <u>my pain could not be more severe</u>. Patients did not have access to prior scores. The VAS data was treated in a millimeter grid, while the four-point scale was scored from one to four. Statistical analyses were performed both on the raw score and on their arcsin transformations to normalize the score distribution. Results indicated that patients preferred the VAS to the four-point scale because they perceived it was more accurate and more sensitive, with better indices of pain. Twenty out of 52 patients preferred the visual analogue scale, although the VAS method of evaluating pain intensity was no more difficult than the interval scale for the patients to understand and complete if they were properly instructed. The VAS was as reliable and more sensitive than the four-point scale in registering the intensity of chronic pain. It appeared that the additional work required to analyze the visual analogue scale would pay off.

Beck Depression Inventory

The Beck Depression Inventory (BDI), the copyright for which did not permit inclusion of a scale, was a 21-item self-report scale designed to measure the severity of depression. Self-administered by paper and pencil, it involved five to ten minutes to complete and required no training. The BDI assessed a state for the day on which it was completed; it was not a trait measure. Scale items for the original depression inventory were derived from

clinical observations and symptoms (Beck, Ward, Mendelson, Mock & Erbaugh, 1961). Later called the Beck Depression Inventory, it was revised in 1974 and copyrighted in 1978. The Beck Depression Inventory is popular for assessing the severity of depression in psychiatrically diagnosed patients as well as in normal populations. According to the Flesch score (Beck, Steer & Garbin, 1988), the reading level of the BDI was reported at the sixth-grade level. Beck, Steer and Garbin (1988) provided guidelines for BDI cutoff scores with patients diagnosed as having an affective disorder: normal or minimal depression at less than 10; mild to moderate depression, 10-18; moderate to severe depression, 19-29, and severe depression at 30--63. Appropriateness of various cutoff score ranges for the BDI varied depending on the sample and the purpose. For example, among university students, high scores on the BDI were not necessarily nosologic for depressive disorder since the "high BDI total scores may just represent diffuse maladaptive functioning in subclinical populations" (Beck, Steer & Garbin, 1988, p. 80). A higher cutoff depression score should be used to minimize the prevalence of false positives. On the other hand, a lower cutoff score would be used to maximize the number of depressed cases. A manual for using the Beck Depression Inventory is available for purchase from the Psychological Corporation.

In the original validation study on the depression inventory, Beck et al. (1961) used a random sample of patients from the psychiatric outpatient department and psychiatric inpatient service

of a metropolitan hospital. There were two patient samples; the original group was comprised of 226 patients, and the other group, 183 patients. The diagnostic subgroups were schizophrenic reaction (28.2%), psychoneurotic depressive reaction (25.3%), and anxiety reaction (1.5%). Patients with organic brain damage and mental deficiency were excluded from the study. Four psychiatrists participated in the diagnostic study using double assessments for each patient (two psychiatrists interviewed each patient while the other two observed through a one-way screen). For depression diagnosis, all psychiatrists used the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (cited in Beck et al., 1961). The psychiatrists also used a four-point scale of none, mild, moderate, and severe for rating depression. Results of the study showed that agreement among the psychiatrists regarding the major diagnostic categories was 73%. The percent of agreement on depth of depression was 97%. Internal consistency for the instrument was analyzed by comparing the score for each of the 21 items with the total score on the depression inventory for each patient. All categories had a significant relationship for the depression inventory. Split-half reliability showed a Pearson correlation between odd and even categories to be $r = .86$. Spearman-Brown correction was $r = .93$. A Pearson biserial r was computed to determine the degree of correlation between the scores on the Depression Inventory and the clinical judgment of "depth of depression." Study I with 226 subjects

showed r = .65, p < .01; Study II (n = 183) showed r = .67, p < .01. Beck et al. also analyzed whether the Depression Inventory distinguished groups. Data in Study I were analyzed and cutting scores were established. The same cutting scores were used for Study II. In Study I, the cutting score discriminated between those two categories in 73 out of 83 cases (88%) and, in Study II, in 59 out of 65 cases (91%).

The investigators also assessed the Depression Inventory's ability to assess changes after a time interval. The time interval ranged from two to four weeks. The depression inventory scores changed in all cases when there was a change from one "depth of depression" category to another as assessed by a psychiatrist. To assess criterion or concurrent validity, the investigators identified that they needed another standard against which the Depression Inventory could be judged. They had the diagnosticians formulate judgments of the intensity of depression. They found a high degree of consistency among the psychiatrists' ratings. The investigators concluded it could not be assumed that clinical evaluation was the ultimate criterion, but it was the best one available at the time of the study.

In a final note concerning application of the Depression Inventory, several advantages were cited. Not only was it was more economical than a clinical psychiatric interview, but it provided numerical scores, reflected changes in the depth of depression over time, and provided measures for judging clinical improvement.

However, the investigators pointed out that the instrument was designed to assess varying degrees of depression on a continuum. It was not designed to differentiate different diagnostic categories; for example, depression may still occur in two persons with different psychiatric diagnoses.

Based on factor analysis (Beck & Beamesderfer, 1974), three factors were extracted: negative view of self and future, physiological symptomatology, and physical withdrawal. Interrater reliability was not done since it was a self-report. Pearson *r* (split-half) was reported to be .86.

Based on a content analysis by Lambert, Christensen, and DeJulio (1983), the Beck Depression Inventory tapped six of the nine symptoms listed in the Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association (APA). Items 18 and 19 of the Beck Depression Inventory assessed decreased appetite and weight but there were no items for increases in appetite, sleep, and psychomotor agitation or retardation. Beck, Steer, and Garbin (1988) argued that there was rationale for the absence of questions about increased appetite, sleep, and agitation. Steer and Beck (1985) did not agree that the scale should have additional questions about increased weight, appetite, sleep need, and agitation to make it more congruent with the APA's Diagnostic and Statistical Manual. Beck (cited in Beck, Steer, & Garbin, 1988) reported the occurrence of loss of appetite at 72%, and the occurrence of sleep disturbance at 87% in severely

depressed patients. Beck, Steer, and Garbin argued that, because increased appetite and sleep occurred frequently in normal population groups, this could produce a high rate of false positives, and agitation was not appropriate to measure in a self-report scale. Lastly, the BDI was to be used for the purpose of measuring intensity of depression in patients with psychiatric diagnosis; it was not developed to be a screening tool. Questions that were atypical of depression were excluded (Beck & Steer, 1985), although the current DSM-III-R included these atypical symptoms as part of the diagnostic criteria.

The BDI was the instrument of choice for sensitive measure of drug treatment, as it measured consistently across inpatients and outpatients (Lambert, Christensen & DeJulio, 1983).

The BDI's validity in measuring depression was further supported by Hill and Kemp-Wheeler (1986) who conducted a factor analysis using both varimax and orthogonal rotations. Varimax factor analysis of the BDI items yielded seven factors from a sample of 160 students which accounted for 41.3% of total variable variance. Six factors were extracted for a sample of 65 psychiatric patients which accounted for 48.7% of total variance. Student and patient factors were almost identical except for one factor reflecting "retardation" or "loss of vital energy". The student data did not yield a factor which could be given this interpretation. The investigators concluded that, overall, the BDI

factor structure in the student sample was comparable to that in the patient sample.

There was found to be a significant negative relationship, $r = -.55$, between BDI scores and social desirability as measured by a modification of the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, cited in Beck, Steer & Garbin, 1988). This would argue for weak construct validity. On the other hand, depressed persons often had low self-esteem and negative attitudes which were expected to be related to the perception of self as undesirable (Beck, Steer & Garbin, 1988). Factor analysis revealed other highly interrelated factors reflecting negative attitudes, performance difficulties, and somatic complaints.

There were several issues and recommendations regarding application of the BDI (Kendall, Hollon, Beck, Hammen & Ingram, 1987). Depression can be either a symptom or a disorder. Further, depression can be a primary disorder in which it has a nosologic category of its own with specific course, prognosis, and treatment, as in diabetes. On the other hand, a diabetic, for example, may have depressed symptoms which do not constitute the disorder of primary depressive illness. The BDI was designed as a "sensitive measure of syndrome depression, but it was never intended to be a nosologic screening device" (Kendall et al., 1987, p. 290). One concern for the BDI was the specificity of the scale for assessing depression as a distinct pathological entity (that is, disease). Research studies found that college students can earn high scores

but not have the primary depressive disorder. Therefore, "the BDI can be viewed favorably as a measure of syndromal depression, but BDI scores alone are insufficient as indices of nosologic depression" (p. 292). Kendall, et al. prudently suggested: assessments of depression should be repeated with multiple screenings to reduce false positives; multiple method assessment should be conducted before imposing the nosologic category of depression on an individual, and the term "depression" should be reserved for individuals with BDI scores over 20 and diagnoses confirmed by structured clinical interviews.

In a 25-year review of psychometric properties of the BDI, a meta-analysis of its internal consistency was conducted by Beck, Steer and Garbin (1988). Results yielded a mean coefficient alpha of .86 for psychiatric patients and .81 for nonpsychiatric subjects. Concurrent validity of the BDI (mean correlations) with clinical ratings was .72 and .73 for the Hamilton Psychiatric Rating Scale for Depression (HRSD; Hamilton, cited in Beck, Steer and Garbin, 1988), an observer rated scale. For nonpsychiatric subjects, the mean correlations were .60 with clinical ratings and .74 with the HRSD. There was evidence that the BDI discriminated subtypes of depression and distinguished anxiety from depression.

Leserman and Koch (1993) criticized the BDI on one issue: they believed that the measure was more demanding of subjects than inventories using Likert-type items, since each of the four response categories was different.

Zung Depression Scale

Zung (1965) identified a need for assessing depression based upon several inadequacies of scales: length of questionnaire, time consumed, and skill of the interviewer. Wanting a scale that clients could complete based upon their own responses, was easy to quantify, and rated severity of depression, Zung devised a 20-item scale which was intended to measure depression as a disorder on a four-point Likert scale with the following anchors: a little of the time, some of the time, good part of the time, most of the time (p. 65). The items were generated based on common characteristics of depression that were reported in the literature (Grinker, 1961; Friedman, 1963) and clustered these characteristics into three factors: "pervasive affect, physiological equivalents or concomitants, and psychological concomitants" (Zung, 1965, p. 63). Following this step, Zung developed the scale based on these factors and verbatim records from patient interviews. Subsequently, there were three items in the Zung Self-rating Depression Scale (SDS) that were not listed in the APA's Diagnostic and Statistical Manual: diurnal (daily) variation, constipation, and tachycardia. The final scale was devised to contain a total of 20 items, 10 of which were worded negatively and 10, positively. The 20 items were each rated on a scale of 0-4, giving a maximum possible score of 80; scores were expressed as a raw score. Each item was rated according to frequency of occurrence rather than intensity of the symptoms. An index for the SDS was obtained by

dividing the sum of the raw scores obtained on the 20 items by the maximum possible score of 80 and expressed as a decimal: below 50, within normal range; 50-59, minimal to mild depression; 60-69, moderate to marked depression; 70 and over, presence of severe to extreme depression.

Zung administered the SDS to all patients who were admitted during a five-month period to the psychiatric service of a hospital and who were diagnosed with depression on admission. Fifty-six patients were tested and admitted with depression. However, 25 of the 56 patients were found to have other psychiatric diagnoses after admission; they were not treated with antidepressant drugs or electric convulsive therapy. The SDS was also given to a normal control group of 100 individuals who were free of observable symptoms and who had no history of recent depressive illnesses. Depressed persons ranged from .63 to .90 with a mean of .74 (corresponded to a raw score of 59), which was significantly different from the controls which ranged from .25 to .43 with a mean of .33 (corresponded to a raw score of 26). The lower the score indicated lower depression; whereas, the higher score indicated greater depression. Reliability by Zung (1972) revealed split-half reliability, $r = .73$, for a psychiatric population. There was no test-retest reliability reported. Hedlung and Vieweg (1979) reported that test-retest was questionable due to the fluctuating nature of depression and recommended that test-retest

reliability be done at two-to-three day intervals instead of the acceptable two to three week interval.

The SDS was validated against an observer-rated Hamilton Depression Rating Scale (HDRS). Validation studies found a correlation of .62 with the HDRS, .73 with the BDI, and .62 with the Visual Analog Scale (Davies, Burrows & Poynton, 1975). The validity of the SDS was questioned by Carroll, Fielding and Blashki (1973) who found that the SDS failed to discriminate between or among patients with depression in general practice, day patients and inpatients and found that the correlation of the SDS with the HDRS was only .41.

There were other problems with the SDS. Based on the investigator's content analysis, there were criteria from the APA's Diagnostic and Statistical Manual that were not included as items on the SDS; these were: increased appetite, psychomotor retardation, guilty feelings, and hypersomnia. Other problems reported by Lambert et al. (1983) included: (a) it did not differentiate among psychiatric groups or at levels of depressive symptomatology; (b) it was not well researched as a psychotherapy measure or as a drug measure, and (c) it was least sensitive to treatment effects when compared to other instruments. They concluded that "it is difficult to defend the use of the SDS as an outcome tool in treatment outcome studies" (p. 267).

Levine-Pilowsky Depression Questionnaire

To develop the Levine-Pilowsky Depression questionnaire, 57 yes versus no checklist items were developed from textbook descriptions of depression. In a validation study, Pilowksy and Spalding (1972) found that 25 of the items discriminated between depressive and nondepressive groups. Concurrent measures used to validate the scale were the visual analogue scale, one of which was self-rated and the other observer-rated. Results showed low correlations: self-rating, $r = .59$; observer rating, $r = .62$.

Wakefield Self-assessment Depression Inventory

The Wakefield Self-assessment Depression Inventory was a 10-item self-report depression scale developed by Snaith, Ahmed, and Hamilton (1971), who evaluated the psychometric properties of the scale and found that 3% of normals and 7.5% of depressives were misclassified. Correlation with the Hamilton Depression Rating Scale was .89. Test-retest correlations were low at .68. The rating scale was compared against clinical global rating and found to be weak in differentiating levels of severity. It was recommended that it be dropped from use (Thompson, 1989).

The Center for Epidemiologic Studies Depression Scale

Developed by Radloff (1977) as a short (15 minutes to administer), 20-item self-report scale, the Center for Epidemiologic Studies Depression (CES-D) scale was designed to measure depressive symptomatology in the general population for

epidemiological studies. Respondents were asked how often over the past week they had experienced each of the 20 symptoms in the CES-D. Responses were rated on a four-point scale: (0) <u>rarely or none of the time</u> (one day a week); (1) <u>some or a little of the time</u> (1-2 days a week); (2) <u>occasionally or a moderate amount of the time</u> (3-4 days a week), and (3) <u>most or all of the time</u> (5-7 days a week). A total sum of the responses became the scale score. The possible range of scores was 0 to 10, with the higher scores indicating more symptoms of depression, weighted by frequency of occurrence during the previous week. The purpose of the CES-D scale was different from other scales which measured for severity of illness or for diagnosis and evaluation. The CES-D was developed for use in studies of the epidemiology of depression among the general population, and it measured the current level of depression. Accordingly, it assessed prevalence of depression not severity or diagnosis. The CES-D could be administered either by self-report or interview procedures.

The scale items were developed from a pool of items of previously validated depression scales such as the Beck Depression Inventory and Zung's Self-rating Depression Scale, as well as from the literature and factor analysis studies. Item components included: depressed mood, feelings of guilt and worthlessness, feelings of helplessness and hopelessness, psychomotor retardation, loss of appetite, and sleep disturbance; for example, one item was "I was bothered by things that usually don't bother me."

Two probability samples of households were selected in a validation study by Radloff (1977). One person aged 18 or over was randomly selected for interview from each household in the sample. Demographic characteristics were not given for this study, although the investigators reported that the samples were probably underrepresented for males and for the poorly educated. A psychiatric patient sample was used which was composed of 70 patients who were residing in a private psychiatric facility and who were selected on the basis of willingness and ability to participate. Although the scale was a self-report, there were differences in the data collection procedure. Initial scores were based on interviews, yet the test-retest data was collected by self-administration. This diversity in data collection for an instrument that was designed to be a self-report was a limitation of this study. In reference to the two samples, more heterogeneity was expected in a general population sample than with the patient sample. The expected result was lower inter-item correlations, but the direction of correlations should still be consistent enough to produce high measures of consistency. For the patient group, expected results were: higher item means, higher inter-item correlations, and very high internal consistency. The results supported these expectations. Coefficient alpha and Spearman-Brown was .85 for the general population and .90 for the patient sample. Test-retest correlations were moderate ranging between .45 and .70. The CES-D scores discriminated well between psychiatric inpatient

and general population samples and discriminated moderately among levels of severity within patient groups. The CES-D score for the group of 70 psychiatric inpatients was substantially and significantly higher than the average for the general population sample. The CES-D had the highest correlation with other scales designed to measure depression in the patient samples. For the patient sample, the highest correlation was .70 with the adjective checklist for depression (Lubin, 1965).

The scale had some limitations for use, as it was not meant to be a clinical diagnostic tool. Cut-off scores for clinical screening were not validated. There was also some question as to the effect of the interviewer and the interview form on contaminating scale scores.

Coyle (1991) examined psychometric properties of the CES-D scale among 790 adults with a physical disability using a structured personal interview that was designed to study the relationships between leisure, work, and life satisfaction. The CES-D scale was administered, and it was reported that adults with a physical disability related the persistence of depressive symptoms significantly more often that did non-disabled adults. Factor structure of the CES-D scale yielded four factors explained 54% of the total variance. These factors were labeled depressed, somatic, positive effect, and interpersonal. The total CES-D scale had a Cronbach alpha of .90.

Carroll Rating Scale

The Carroll Rating Scale (CRS), a self-rating adaptation of the 17-item Hamilton Rating Scale (HRS), was an interview/rater scale designed to measure severity of depression (Carroll, Feinberg, Smouse, Rawson, & Greden, 1981). The CRS was comprised of 17 items which were non-dichotomously scored on a four-point (0-3) scale according to four statements. A convenience sample of 119 adults aged 18 to 64 who were employed in a university setting was used to test the CRS. For item analysis, item correlations on the correlation matrix ranged from .19 to .78 with a median of .54. The researchers inferred content validity from content analysis of the CRS with the Hamilton Rating Scale. Concurrent validity was estimated by comparing CRS scores with HRS scores in patients with depression. Results showed the matrix of correlations between items and total scores for 278 matched HRS-CRS ratings in 97 patients with depression was $r = .80$. Factor analysis revealed severity of depression and anxiety. There was no Cronbach's alpha reported for a coefficient of internal consistency. Test-retest was not performed. Results were indicative of a good pretest, but much more evidence was needed to justify its use in different settings. As with the Beck Depression Inventory, there were problems with the CRS, as it was not in complete concordance with the DSM since it was missing items on hypersomnia and appetite changes (the descriptor for "decreased" was included but not "increased"). Laserman and Koch (1993) concluded that the CRS

needed more studies to confirm the results and to investigate the sensitivity of the CRS to psychiatric treatment.

Inventory to Diagnose Depression

The Inventory to Diagnose Depression (IDD) was a 22-item self-report scale designed to diagnose major depressive disorder utilizing 1980 criteria of the Diagnostic and Statistical Manual. In a validation study, Zimmerman and Coryell (1987) reported reliabilities as follows: split-half $r = .91$ and Cronbach's alpha was $r = .92$. The study used the Diagnostic Interview Schedule (DIS) developed by Robins, Helzer, Ratcliff and Seyfriend (cited in Zimmerman & Coryell, 1987), a highly structured interview tool designed for use by lay interviewers in epidemiological studies of psychiatric disorders. Zimmerman and Coryell compared the prevalence estimates of major depressive disorder using the IDD and DIS, as well as concordance between the two instruments. The relation between the overall rate of agreement on the DIS and the IDD was 97.2%. The main criticism of this scale was that studies to assess sensitivity to change in clinical status had not been assessed.

Inventory for Depressive Symptomatology, Self-Report

The Inventory for Depressive Symptomatology, Self-Report (IDS-SR) was designed to measure severity of depression in both inpatients and outpatients (Rush et al., 1986). Items were generated from existing inventories, diagnostic criteria, expert

clinical judgment, and the DSM-III of the American Psychiatric
Association (cited in Rush et al. (1986).

The scale contained 28 items, each of which was rated from
zero to three with increasing severity; three was the highest score
for each item. Reliability of the IDS-SR was assessed from item-
total correlations and Cronbach's coefficient alpha which was .85.
Items found to correlate the least to the total scores were:
distinct quality of mood, hypersomnia, diurnal variation of mood,
mid-insomnia and early-insomnia, weight change, and somatic and
gastrointestinal complaints. The IDS-SR correlated with the
Hamilton Rating Depression Scale ($r = .67$) and with the Beck
Depression Inventory ($r = .78$), which indicated a marginally
acceptable correlation with the former and fairly good correlation
with the latter. Discriminant analysis revealed that the scale
showed differences in normal control and depressed group scores,
but the differences in classification were not "spectacular" (Rush,
et al., 1986, p. 72).

Factor analysis was also used to established construct
validity. Rush et al. (1986) expected four factors to be
extracted: a general factor measuring mood and cognition, a
vegetative (endogenous) factor, an atypical symptom factor (weight
gain, hypersomnia, appetite increase), and an anxious depression
factor. Setting the number of factors to be extracted to four,
results showed that the four factors accounted for a cumulative
variance of 45.1%.

Validation of this scale was limited by the relatively small sample size of 23 normals (persons who were free from depression and medical problems, who were not taking medications, and who had a negative family history of psychiatric problems). Based on the investigators' content analysis of the scale, there were no items that captured excessive guilt.

Rush (1993) reported scoring of this scale as follows: normal, less than 12 or less than 15; dysthymics (not major depression), 18 to 28; major depressives, 35 and above.

Depression Check-up

The Depression Check-up (Schiraldi, 1987; Appendix A) was a 22-item, self-rating scale that measured severity of depression and involved about five minutes or less to complete. The scale was developed by Schiraldi (1987, 1990) based on a review of the literature and the DSM-III-R definition of depression. Scoring on the Depression Check-up was based on scale scores from the Beck Depression Inventory. The higher the grand total, the more severe was the depression.

The investigator selected this scale for several reasons. Of primary importance, it met the DSM-III-R criteria for depression. It involved five minutes or less to complete and appeared to have applications in health education, college, and primary care settings for screening purposes. The self-scoring method was unique to self-rating depression scales, as a self-scoring scale is

ideal for use in self-help booklets and publications as an "early
warning sign" or "depression thermometer" for individuals who are
depressed or are at risk for developing a depression. The
Depression Check-up met most all of the investigator's criteria for
a self-rating depression scale. It was different from the Correa-
Barrick Depression Scale (CBDS) since it used a Likert-type format
instead of the visual analog format. The investigator believed
that it was ideal as a screening tool.

Based on the investigator's content analysis of the
Depression Check-up with the DSM-III-R, the scale was in accordance
with the DSM-III-R criteria for depression. The scale was
initially piloted by Schiraldi (1987) on 200 university students to
determine its convergent validity with the Beck Depression
Inventory. Results showed that it positively correlated .80 with
the Beck Depression Inventory.

Observer-rated Scale

Since many of the self-report scales had been validated with
the Hamilton Depression Rating Scale (HDRS), it was useful to
review this scale. The original Hamilton scale (1960) was
developed as a 17-item, observer-rated scale to be completed by a
skilled psychiatrist. It required much clinical skill to complete,
and, therefore, it could not be used by an interdisciplinary team.

Based on a review of validation studies of the HDRS,
interrater reliability was consistently high, ranging from .87 to

.98. Item correlations ranged from .45 to .78, and the scale showed varying sensitivity to change based on correlation studies with scores during treatment. Correlations ranged from a low of .38 to a high of .72. Factor analysis by Hamilton (1967) showed two factors which accounted for almost all the variance: general severity and bipolar with anxiety/agitation at one end and retardation/suicidal ideation at the other.

Scale Summation

In summary, the depression self-rating scores reviewed thus far were either deficient in item content or had not been sufficiently validated. Thus, the Schiraldi Depression Check-up and the Correa-Barrick scales were designed in an attempt to improve upon current, published scales.

Scale Development

Before developing a new measure, key design features such as text, item construction, and format should be considered. Some suggestions for scale development relate to the text of the survey. For example, use of multiple negatives should be avoided because the negative "not" in textual material might be confusing to the reader. In addition, using negatively as well as positively worded items within the same scale avoids acquiescence bias. Consideration also needs to be given to the disadvantages of

reversing items on a scale, as this could cause the scale to be confusing to respondents.

A variety of ways were suggested for formatting a scale. Two of the most popular and most relevant to the scales under study were the Likert Scale (DeVellis, 1991) and the Visual Analogue Scale. Likert scales have frequently been used to assess opinions, beliefs, and attitudes. Item responses are in a declarative format followed by response categories that reflect intervals of agreement or endorsement and which are scaled in graduated intensity along a continuum. The Visual Analog Scale presents the respondent with a continuous line between a pair of descriptors representing opposite ends of a continuum (DeVellis, 1991). A major advantage is its sensitivity which can be helpful when measuring changes before and after an intervention or experimental manipulation. Another advantage is that respondents have a difficult or impossible time memorizing past responses. However, a main disadvantage to the Visual Analog Scale was whether a mark placed at one point on the scale meant the same thing as a mark placed there by another person (DeVellis, 1991).

Providing guidelines for scale development, DeVellis (1991) outlined several steps:

Step One: Determine what is to be measured. Determining what is to be measured involved understanding the theories or conceptual basis of the construct or it meant having a clear definition of the construct in the absence of a theory. A scale containing items at

the same level of specificity or generality with regard to the concept being measured usually increased reliability. To avoid "cross over" of items into a related but different construct, the scale should be developed with a clear sense of what items were irrelevant or superfluous to the concept being measured.

Step Two: Generate an item pool. Generating an item pool involved choosing items that reflected the scale's purpose. Items should be sufficient in numbers to allow the investigator a large enough pool from which to select final items after the item analysis. Also, items can be redundant, thus giving the investigator more choices. Careful attention needs to be given to item wording, and, generally, lengthy items should be avoided since this increases complexity and diminishes clarity. Since reading level should also be of interest in scale development, DeVellis recommended that a reading level between the fifth and seventh grades be used for most instruments with a general population. To avoid acquiescence or agreement bias, items should also be worded positively and negatively and should not be worded double-barreled such as "When did you stop beating your spouse?" According to DeVellis, a double-barreled item forces the respondent to either agree or disagree with either or both ideas. Other editorial suggestions for item writing included avoiding ambiguous pronoun references such as "they" and misplacement of pronoun references.

Step Three: Determining the format. Since the scale developer has a variety of formats from which to choose, the format

should be decided upon very early. As a guideline, scores made up of items evaluated and summed on a continuum are desirable. Scale items should be equally weighted so that they are generally parallel. Whether a construct being measured is influenced by time can also be an issue; for example, depression scale scores can be influenced by time. If time is an issue, then the scale should contain explicit directions for the respondent.

Step Four: Have the initial item pool reviewed by experts. The scale developer should have a panel of experts, who have been given an operational definition of the construct, review the scale for relevance, clarity, and consistency. Reviewers can suggest additional items that would enhance the scale's content validity, but the final decision would rest with the investigator.

Step Five: Consider the inclusion of validation items. This involved including items that measured social desirability to assess whether this may be a problem for the instrument. For example, the investigator could include a scale on social desirability to see if it correlated with the scale being evaluated.

Step Six: Administer item to a development sample. The size of the development sample has an effect on internal consistency, as too small a sample can artificially inflate the correlation of the scale items and can create problems with the generalizability of the scale results across populations.

Step Seven: Evaluate the items and Step Eight: Optimize scale length. Both of these steps involve statistical procedures which will be discussed under validity and reliability.

Specific suggestions for scale format were reviewed extensively by Berdie, Anderson and Niebuhr (cited in DeVellis, 1991). In summary, the scale developer should consider several important procedures: (a) begin with non-threatening questions; (b) group items into logical sections; (c) do not put items at the end of a questionnaire; (d) provide transitions between sections; (e) number items to avoid confusion; (f) put the study title in bold type on the first page of the questionnaire; (g) design questionnaire attractively; (h) use brief instructions; (i) use "white space," and (j) avoid using the words questionnaire of "checklist" on the form since some people are prejudiced against these words.

Of the variety of methods for formatting a scale, the Visual Analog Scale (VAS) was the focus of the present discussion. The VAS was similar to the semantic differential and was discussed by DeVellis (1991). In use, this format presents respondents with a pair of polarized responses that represent opposite ends of a continuum such as very sad to very happy. The individual marks a point on the line that represents a response for that item. Advantages to this included: (a) it had the potential for being very sensitive, (b) it was difficult or impossible for subjects to memorize their responses, and (c) it could be used to detect

subtle, mild treatment effects which a five-point scale may not demonstrate. On the other hand, one major disadvantage of the visual analog was that a mark placed at one point on the line may have diverse interpretations.

Reading Level

Reading level was an important issue in scale development especially for self-rating scales. The software package "Rightwriter" was used for determining writing style and the reading level. For the Depression Check-up (Schiraldi, 1987), results suggested that readers needed a 12th-grade level of education to read it; the strength of delivery was good but could be improved upon; the use of adjectives and adverbs was normal; there was an absence of jargon, and few compound sentences or subordinate clauses were used.

According to "RightWriter" software, results on the Correa-Barrick Depression Scale indicated that readers needed a second-grade level of education; the writing can be made more direct by using the active voice with fewer weak phrases and more positive wording; the use of adjectives and adverbs was normal; there was an absence of jargon; few compound sentences or subordinate clauses were used; many sentences started with adverbs, and few prepositional phrases were used. Suggestions from "Rightwriter" were used to edit the Correa-Barrick Depression Scale.

Validity and Reliability

The discussion and procedures in this study for establishing scale validity and reliability were based on classical test theory. Crocker and Algina (1986) described test theory as "a general framework for viewing the process of instrument development" and "the study of the pervasive measurement problems . . . and methods for their resolution" (p. 7).

There were five problems commonly encountered in developing measurements of psychological constructs:

1. No single way of defining a psychological construct was universally accepted.

2. Psychological measurements were based on samples of behavior.

3. Sampling of behavior resulted in errors of measurement.

4. The units of measurement were not well defined.

5. The measurements must have demonstrated relationships to other variables to have meaning (Crocker & Algina, 1986, p. 13).

Validity

According to the American Psychological Association (APA), validity was considered the most important factor in test validation, and it referred to "the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores" (1985, p. 9). Validity was defined as "the assurance that results obtained from measurement or evaluation are an

accurate reflection of reality (Green & Lewis, 1986, p. 367). Ideal validation included several types of evidence, but the quality of evidence was of greater importance than the quantity of questionable evidence (APA, 1985).

There were several procedures for validity. While face content referred to whether the scale appeared to measure what it was supposed to measure, content validity entailed the following steps: (a) defining the performance domain of interest, (b) selecting a panel of qualified experts in the content domain, (c) providing a structured framework for the process of matching items to the performance domain, and (d) collecting and summarizing data from the matching process (Crocker & Algina, 1986, p. 218).

There were several statistical tests for establishing validity, one of which was criterion-related validation. The design of a criterion-related validation study should have the following steps:

1. Identify a suitable criterion behavior and a method for measuring it.

2. Identify an appropriate sample of examinees representative of those for whom the test will ultimately be used.

3. Administer the test and keep a record of each examinee's scores.

4. When criterion data are available, obtain a measure of performance on the criterion for each examinee, and

5. Determine the strength of the relationship between test scores and criterion performance (Crocker & Algina, 1986, p. 224).

Issues in content validity to be considered were: (a) instructions to the panel, (b) degree to which the items represented the construct, (c) meaningfulness of the items to different cultures, and (d) relevancy of item performance data (from the item analyses) to the judgement of content validity.

Crocker and Algina (1986) differentiated predictive and concurrent validity. Predictive validity referred to the "degree to which test scores predict criterion measurements that will be made at some point in the future" (p. 224); for example, Scholastic Aptitude Test scores can predict a college grade point average. Concurrent validity referred to "the relationship between test scores and criterion measurements made at the time the test was given" (p. 224). It was noted that several problems could contaminate a criterion validation study; for example, selection of an appropriate and feasible criterion, inadequacy of the sample size, and lack of reliability of the predictor or criterion measure. Concurrent validity was usually reported as a validity coefficient.

Construct Validity

Construct validity was defined as "the degree to which an instrument measures the construct or trait it was designed to measure" (Wilson, 1985, p. 564). Procedures for construct

validation may include (a) correlations between test scores and designated criterion variables, (b) differentiation between groups, (c) factor analysis, and (d) the multitrait-multimethod matrix analysis (Crocker & Algina, 1986; Campbell & Fiske, 1959). The latter method involved convergent and divergent validity. Convergent validity was the correlation between measures of the same construct using different instruments. Divergent validity (also referred to as discriminant validity and heterotrait-heteromethod) was the correlation between different constructs using different measures; the correlation should be lower.

Factor Analysis

Polit and Hungler (1983) defined factor analysis as a multivariate procedure that involved a higher degree of subjectivity than most statistical procedures because it did not test hypotheses. The purpose of factor analysis was to reduce a large set of variables into smaller, more manageable sets, and it was used as a procedure to determine the number of factors which were unobservable or latent variables or clusters of variables in a measure. In other words, the process of factor analyses identified latent variables that can account for the covariances of items among them. The results of factor analyses could be used to condense the scale so that fewer items were needed. Factor analysis could help an investigator examine how much of the total variance could be accounted for as additional factors were added.

There were separate stages to factor analyses, the first of which was factor extraction. The most popular method for initial factor extraction was "principal component" or "principal axes"; this method produced a factor matrix which showed coefficients for each variable. In factor analysis, a decision has to be made regarding the number of factors indicated by the correlation matrix. There were many criteria for determining the number of factors; one was the number of eigenvalues of the correlation matrix that was greater than 1.00 (Crocker & Algina, 1986). Polit and Hungler (1983) defined eigenvalues as values equal to the sum of the squared weights for each factor. A decision about the number of factors to retain was based upon the "eigenvalue rule" by Nunnally (1978): only factors which explained more variance than the average amount explained by one of the original items should be retained. If one or more factors were explaining less variance than an item, then they should not be retained. In general, eigenvalues greater than 1.00 suggested a factor.

The second method for determining the number of factors was the scree test: a vertical axis corresponding to eigenvalues, a horizontal axis corresponding to successive factors, and numerical markers plotted on these axes indicating the eigenvalues corresponding to each factor. When values are plotted graphically from the top left of a graph to the bottom of the graph, there will be a point on the graph at which an "elbow" occurs; this is the junction at which a factor (or factors) diminishes its ability to

explain variance. The portion below the elbow is called the scree; factors above the scree are retained and those below are not (Kim & Mueller, 1978).

As the factor matrix produced by an unrotated factor extraction is difficult to interpret, factor rotation is performed on those factors that have met one or more of the criteria for inclusion. In factor rotation, there are two types of rotation which are geometric representations: orthogonal solutions (also called Varimax) which are uncorrelated factors and oblique (direct quartimin) solutions which are correlated factors. Solutions for the number of factors are determined by plotting variables so that they cluster within new horizontal and vertical axes. For orthogonal solution, the new axes are perpendicular to one another. In oblique, they are not. Variable factor loadings are defined as regression weights, and their loadings on a rotated axis are examined. Factor loadings less than .30 are usually considered unimportant. The best test of the number of factors is replicability of the factor analysis (Crocker & Algina, 1986).

Finally, the number of factors may vary with the number of factor solutions one wishes to interpret. For example, one variable may define a factor, but some analysts would argue that one variable to explain one factor is uninterpretable.

Reliability

    For the following discussion, the terms "scale" and "test" are used interchangeably.

    According to Crocker and Algina (1986), "desired consistency or reproducibility of test scores is called reliability" (p. 105). The following discussion on reliability procedures is based on the classical true score model which viewed each observed score as the sum of the examinee's true score and a random error component. The true score is "the expected value of the examinee's test scores over many repeated testings with the same test" (Crocker & Algina, 1986, p. 127).

    One method for establishing reliability is to assess the coefficient of stability which is the correlation coefficient obtained from a test-retest procedure. In this procedure, the investigator administers a test, waits, and then readministers the same test to the same group; the investigator then computes the correlation coefficient between the two sets of scores (Crocker & Algina, 1986). There is no fixed rule for the time period for when a test should be readministered. A general guide is that the test should be readministered after a long enough time period so that respondents cannot recall their prior responses but not so long that historical or maturational issues pose a threat. A major issue with test-retest is whether the coefficients are influenced by other factors such as test sensitization and practice. For this

reason, it is prudent to employ another reliability method which uses a single test administration.

Reliability may also be verified using a single test which is termed the test's internal consistency. This can be estimated by split-half method wherein the scale developer administers one form of the test but divides the items into two equal subsets. The easiest way to divide items on a scale is by either random assignment of items to the two half-test forms or assigning odd-numbered items to one form and the even-numbered to the second form. One problem with this procedure is that the reliability coefficient is likely to be lower for the split-half test than for the full-length test. The Spearman Brown formula can be employed for correction, but this statistic assumes that the half-tests are strictly parallel. Tests are parallel when each examinee has the same true score on both forms of the test, and the error variances for the two forms are equal and matched in content. Such tests will have equal means and equal variances.

There are several popular procedures for estimating reliability coefficients: Cronbach's alpha, Kuder Richardson (KR) 20, and Hoyt's analysis of variance. All of these procedures produce the same results by coefficient alpha, which is "the average of all the split-half correlations that would be obtained if the test were divided into all possible half-test combinations using Rulon's procedure" (Crocker & Algina, 1986, p. 153). Cronbach (1951) developed coefficient alpha for use with a single

test administration, and it can be estimated from items that are scored either dichotomously or non-dichotomously. Kuder-Richardson (KR) 20 can only be used for dichotomous items. Hoyt's method is based on analysis of variance and is an alternative method for estimating reliability without computing alpha.

Reliability coefficients are affected by several testing situation issues. If the sample is very homogeneous on the trait being measured, the reliability estimate will be lower than if it were heterogenous. Longer tests are usually more reliable than shorter ones.

Effects of decreasing or increasing the number of items in a scale on the alpha level can be determined by the Spearman Brown formula; however, this procedures assumes that the items are parallel in content to the original.

Item analysis, another procedure for estimating internal consistency, was defined as "the computation and examination of any statistical property of an item response distribution" (Crocker & Algina, 1986, p. 335). A correlation matrix is produced and then inspected to determine if each scale item is intercorrelated. The higher the correlations among the scale items, the higher the reliability. The correlation matrix is examined for item variances and item means. It is desirable for the scale items to possess high variance, which reflects good discrimination, and a mean close to the center of scale range (Crocker & Algina, 1986). For interval data, the correlation matrix is composed from the Pearson

product moment correlation. Item analysis should be done on five to ten times as many subjects as items (Nunnally, 1967); in other words, a 20-item test, should be administered to at least 100 subjects. Information from item analyses should be used for revising or deleting items based on an inspection of negative correlations.

One canon in classical test theory is that the reliability coefficients also reflect the characteristics of the sample being tested as well as the test. This means that an interpretation of reliability must always be made within the context of the sample.

Measurement Error

There are two types of errors to be concerned about in validation: random and systematic. Random error occurs by such chance happenings as distractions during the test or administration or scoring errors. Systematic error occurs when certain characteristics of the respondent which affect the score have nothing to do with the construct being measured (Crocker & Algina, 1986). In the first type, social desirability is used to describe a biasing factor that results from the respondent's attempt to demonstrate behavior that is socially desirable or preferred (Green & Lewis, 1986). In the second, acquiescent response set results from the respondent choosing a particular pattern of behavior that does not vary in response to the measurement stimulus. For example, a respondent may select all extreme answers, may zigzag

across response options, or may choose only neutral response options. Valid instruments minimize the amount of systematic error.

The researcher should examine the degree of "random error" or the discrepancy between a respondent's true score and observed score over repeated testings which is called "the standard error of measurement" (Crocker & Algina, 1986). The standard error of measurement is an interpretation requiring confidence intervals in which the score is expected to fall. The standard error of measurement should be reported for each reliability estimate. When a test is administered to different populations, the standard error or measurement should be reported for each sample.

The question frequently asked is: How high does the reliability coefficient have to be in order for the scale to be considered reliable (Beck, 1989). Beck believed that a researcher should strive for as high a reliability as possible, but often scales can possess acceptable levels of reliability ranging in the upper 0.70s.

Types of Psychotherapy

A review of the literature on strategies for intervening with depression, especially related to the discipline of psychology, was relevant since it framed the stage for analyzing implications for the health education role. Since psychologists, counselors, and health educators do not prescribe medication, it was beyond the

scope of this section to review pharmacological interventions. However, the reader is cautioned to recognize its therapeutic value.

There were different types of psychotherapy, and each was reviewed in terms of goal, method, and effectiveness. Psychotherapy has a pivotal role in treating depressed patients; in combination with medication, it was found to be very effective. On the other hand, for some patients who have a depressed mood without the diagnostic features of a major depression disorder, psychotherapy was sometimes found to be all that was needed.

Educative Therapy

The concept of education as one healing tool was introduced as a model for health education practice. Educational programs can help individuals to learn about the early warning signs of a major depression, to seek early treatment, and to adopt behaviors to inoculate against the stressors which may trigger the depression.

The idea that education was therapeutic has appealed to health educators and has received support in the literature (Authier, Bustafson, Guerney, & Kasdorf, 1975; Kaminsky, 1986; DePaulo & Ablow, 1989). Authier et al. discussed the role of education as a therapeutic modality: "the person being served is seen as analogous to a pupil, rather than a patient" (p. 31). Authier et al. felt that the bio-psychiatric approach was the most appropriate treatment and elaborated on the idea of combining the

medical and educational models of treatment of psychiatric patients.

Kaminsky (1986) utilized an educative-supportive technique that viewed affective disorder as a disease that constrained patients to think of themselves negatively and that "the negative thinking of the patient makes sense to him" (Kaminsky, 1986, p. 4). This psychotherapeutic approach included describing to the patient that a major depression was biological and that its resolution was not within the patient's control. The goal was to provide patients with a cognitive tool to help them combat their change in self-attitude. Although educational in nature, this method mirrored cognitive reappraisal as a stress-control technique. Kaminisky strongly asserted the view on treating major depression with combined pharmacology and psychotherapy:

> The psychotherapy literature concentrates on etiological explanations and all literature, even on the combined use of medication and psychotherapy, used models of depression that assert its psychological origin, whether that be dynamic, behavioral or cognitive. The authors who associate the disease perspective do not address psychotherapy and would leave the impression that all they do for the patient is push pills. (1986, p. 6)

Kaminsky provided a cognitively based tool to help individuals combat changes in self-attitude. This method is focussed on the distortions of patient beliefs during the illness.

DePaulo & Ablow (1990) described supportive therapy as "It seeks to convey the realities about depression or mania to patients whose abilities to understand their illnesses are impaired" (p.

134). To this end, information should be provided to the patient and family about diagnosis, cause, rationale, and prognosis for treatment. Emphasis should also be centered on providing hope to the patient that the depression will lift.

Cognitive Therapy

Cognitive therapy was shown to be effective in depression, especially when used in conjunction with medication. This type of therapy aimed to improve the patient's negative outlet and to decrease feelings of depression, with the therapist confronting the patient with objective, realistic data on the patient's distorted thinking. Some therapists believed the patient's negative thinking was the cause of major depression. Others believed that the negative thinking was a symptom of the disorder. Empirical evidence suggested that it was an effective treatment (Everly, 1990).

Behavioral Therapy

There was no evidence that behavioral therapy was effective for clinically depressed patients. It was useful for other disorders, such as phobias, obsessive-compulsive symptoms, eating disorders, and behavioral problems that may have perpetuated the depression (DePaulo & Ablow 1989).

Implications for Health Education

A core function of health education has been practice. Health education programs have been developed around themes of self-therapy and stress-management. Health educators have helped to prevent and/or lessen the impact of depression by attenuating environmental stressors through the implementation and evaluation of stress management programs. Program content can range from time management to assertiveness training, interpersonal skills development, and relaxation techniques, to name a few.

Stress Management

A stress management program for intervening with a depressed population group would require multiple strategies. Therefore, it was assumed that the best treatment for depressed patients consisted of pharmacological as well as psychotherapeutic interventions. These psychotherapeutic interventions were cognitive, self-treatment, and supportive/educative.

Bandura, Taylor, Williams, Mefford and Barchas (1985) identified the importance of the perception of control in the recovery process, while Beck (1976) identified the importance of a person's interpretation of an event. To this end, cognitive psychotherapy has been used as a tool to alter one's appraisal of an event. There were primarily two types of models. In the first, Ellis (1973) utilized rational-emotive psychotherapy to dispute a

person's irrational beliefs. The following model diagrams this
framework:

A_____B_____C
activating experience   belief   emotional consequence

The core of Ellis's approach was in disputing the clinically
depressed patient's irrational beliefs. Since depressed persons
have negative thinking, Ellis's model was a therapeutic tool for
health education in disputing that type thinking.

    The second model, self-therapy (self-treatment) programs,
relied on patient involvement in eliciting improvement. There was
evidence that self-treatment program subjects showed greater
reduction in depression (Teuting & Koslow, 1983). In self-
treatment, there existed a component of control. Everly (1990)
described this concept as:

> Control has been operationalized as the ability to change an
> environmental transaction, the perceived ability to do so,
> the ability to predict environmental transactions, the
> ability to understand those transactions and/or the ability
> to accept such transactions within some meaningful cognitive
> framework or belief system. (p. 136)

Health educators can promote an individual's sense of control by
(a) suggesting reference books on depression for the patient to
read, (b) promoting optimal health behaviors, and (c) encouraging
self-assessment of depression using self-rating depression scales.

Screening

It is known that one important role for the health educator is screening individuals for depression at health promotion or wellness programs in schools, corporations, and communities. Effective treatment cannot begin until an evaluation has been initiated to confirm or rule out a clinical depression. It is not the role of the health educator to evaluate and diagnose depression but to screen for it. A self-rating tool may be used by the health educator for the purpose of counselling a client that depression is treatable and to obtain further evaluation by a mental health professional. It is appropriate for the health educator to recognize the signs of depression and to offer hope that treatment is available. Often the health educator may be the first person to detect depression in individuals in schools or corporate settings. Since the literature review has supported and documented the numbers of undiagnosed cases, health educators can be on the alert with an index of suspicion for depression in undiagnosed and untreated cases.

Self-rating scales of depression may be used by the health educator in screening for depression in a target population. For example, a health counselor in a university setting may use a self-report depression scale to detect the presence of depression in a college student. If scores indicate the presence of depression,

appropriate interventions and/or referral for further evaluation may be promptly initiated.

Another setting or opportunity for assessment and referral is the health fair. Health fairs are health promotion functions designed for the purpose of health education and selected health screenings (i.e., blood pressure, height and weight, vision, and hearing screening) that are conducted in familiar settings such as schools, worksites, and shopping malls. At health fairs, health educators also screen individuals for depression and provide counselling for either follow-up or referral for evaluation.

The future of health care reform demands that health promotion be accessible, and health fairs are one method for health educators to help achieve this vision.

Patient Education

There has been a dearth of published studies on patient education and depression. Although there have been developments in patient and family education approaches to affective disorders (Schiraldi, 1990; Daley, Bowler & Cahalane, 1992), there were few empirical studies to validate the efficacy. Some data supported the power of informal educational interventions for both patients and families (Daley et al., 1992). However, there was no research that addressed outcome evaluation of informal education interventions as part of a multi-modal approach to treatment. According to Daley et al., there has been a shift away from insight

or interpretative therapy to that providing support and education
for individuals with psychiatric illness. There were reasons for
this shift. Research helped to identify multiple biological as
well as psychological and environmental factors involved in the
development and maintenance of psychiatric illness. Providing this
information to patients and families helped to decrease the stigma
associated with having a mental disorder. A second reason was the
trend in mental health care systems toward decreased length of
hospital stays.

A third reason for education was that many illnesses were
chronic conditions with high rates of recurrence. Educational
interventions for patients and families helped to reduce relapse
rates as well as to prepare them better to identify warning signs
of impending relapses so that actions could be taken quickly to
reduce the length or severity of the episode. Evidence has
accumulated that education worked in helping depressed patients and
their families (Daley et al., 1992). Education in psychiatry has
been coined "psychoeducation" (Daley et al., 1992). The goals of
psychoeducation were to reduce symptoms, increase hope, improve
coping skills, provide emotional support, enhance treatment
compliance, and facilitate cognitive, affective, and behavioral
change. Psychoeducation focuses on topics such as causes and
effects and treatment of depression, effects of medication.
Psychoeducation can be adapted to a variety of clinical populations
(adolescents, adults) and can be implemented by an

interdisciplinary team. Lewinsohn (cited in Daley et al.) implemented a program coping with depression that focused on specific self-change skills such as setting goals, developing plans to change behavior, relaxing, increasing pleasant activities, controlling negative or irrational thinking, and developing social skills which focus on assertiveness and friendship development. An outcome evaluation of the programs using a randomized controlled trial, found it significantly reduced self-reported depression during the one-year follow-up period (Clarke & Lewinsohn, cited in Daley et al., 1992).

Payne (1989) found that health education programs seemed to attract small numbers, while programs centering on stress relief attracted larger numbers. Accordingly, Payne successfully recruited and maintained good member attendance at a health education program by linking it with stress relief. The population consisted of 60 people who enrolled in the course and who attended for a variety of reasons (depression, tranquilizer withdrawal problems, high blood pressure, insomnia and domestic crisis, and anxiety following surgery or myocardial infarction). Content included themes on stress relief, positive thinking, assertion training, depression, anxiety, tranquilizer withdrawal, insomnia, migraine, and loneliness. It also covered topics on physical health such as exercise; circulatory disease; nutrition; weight-reduction; smoking cessation; care of the back, muscles and joints; posture, and self-defense. Methods employed during the relaxation

session included tension release, guided imagery, medication, autogenics and breathing.

National Level Health Education

Health educators can intervene at the national level. For example, the National Institute of Mental Health (NIMH) launched an education program, the Depression Awareness, Recognition and Treatment Program, designed to alert the public and health professionals that depressive disorders were common, serious, and treatable (Regier et al., 1988). In developing the program, the NIMH conducted 20 focus groups in nine different geographic regions and surveyed 500 household units. Findings indicated that most people were knowledgeable about depression but did not know about the changes in physical symptoms nor did they believe that medication would be effective. It was found further that most would not seek treatment because of a perceived negative stigma at work and perceived cost of treatment.

Summary

In this chapter, conceptual areas from the literature were reviewed. Major themes consisted of scale design and psychometrics, clinical and epidemiological research on depression, the role of stress in depression, and the potential applications of the study findings to the health education discipline.

In summary, there is a vast body of knowledge related to instrument development and scale validation. Clinical research findings provided a biological perspective on the cause and nature of clinical depression, and stress is an important risk factor in triggering and/or perpetuating depressive illness. The role of the health educator in stress management cannot be undervalued as an adjunct to traditional pharmacological and psychotherapeutic treatment modalities. Epidemiological research identified high-risk groups which should be targeted for screening. Lastly, the role of the health educator in mental health can be valuable from several perspectives: patient "psychoeducation," stress management, community health depression screening programs, and national-level educational program development.

In the next chapter, the methodology for scale validation for the two new scales, the Depression Check-up and the Correa-Barrick Depression Scale, will be discussed.

CHAPTER III

METHODOLOGY

Introduction

The purpose of this chapter is to describe the methodology used to develop the Correa-Barrick Depression Scale (Appendix B) and to validate the Scale and the Depression Check-up (Schiraldi, 1990; Appendix A). Since instrument design for the Depression Check-up had been completed prior to this study but not for the Correa-Barrick Depression Scale, the process used for developing the latter will be described in this chapter.

There were three phases involved in validating the Depression Check-up and the Correa-Barrick Depression Scale. Validation procedures used were replicated for each scale. Phase I involved qualitative assessment of the Depression Check-up and the Correa-Barrick Depression Scale using a focus group interview and expert panel review for face and content validity. Phase II consisted of a pilot study involving 100 students and 16 faculty and staff members at a metropolitan comprehensive university. Phase III, the main study, was conducted on two different sample sets:

1. Sample I was drawn from 1,200 faculty and staff members at a metropolitan comprehensive university.

104

2. Sample II was comprised of 200 depressed individuals undergoing treatment at a private psychiatric outpatient facility.

Data collections for the pilot and the main studies were confidential and anonymous. Participants were instructed not to include their names on the survey.

Instrument Design

The procedure used in developing the Correa-Barrick Depression Scale, a 30-item self-rating depression scale, was based on suggestions for scale development made by DeVellis (1991, pp. 51-90) and which were discussed in the review of literature.

Construction of the Correa-Barrick Depression Scale began with the purpose of developing a self-report inventory that would measure severity of depression and be sensitive to change in clinical status. This investigator and Correa agreed that a new self-rating depression scale should be designed based upon the visual analog format used in the Correa Depression Scale (Correa, 1983), which was adapted from the Zung Self-rating Depression Scale (Zung, 1965). The multiple-item visual analog format used in the Correa-Barrick Depression Scale was unique for a self-rating depression scale.

The investigator decided that the Correa-Barrick Depression Scale should include all items that would reflect the criteria of the revised Diagnostic and Statistical Manual of Mental Disorders, Revised (DSM-III-R; American Psychiatric Association, 1987) and the

current literature on clinical descriptions of depression. It was also decided that a depression scale might be enhanced by adding items on qualitative descriptions of patients' sensory experiences.

Initial validation of the additional items on "sensory experiences" was sought by a review from a panel of experts (Appendix D). Directions were given to the panel members to review the sensory items for validity.

Development of the Correa-Barrick Depression Scale items was based upon (a) a review and conceptual analysis of the literature, (b) criteria of the DSM-III-R, (c) anecdotal reports of the characteristics of depression by persons with depression, (d) interviews with three psychiatric experts on depression, and (e) items from published self-rating scales.

Sources of data for the scale were also derived from the investigator's interviews of a patient and two psychiatrists (R. DePaulo and S. Simpson) to determine item content (personal interviews, February 11, 1993, Johns Hopkins Hospital). Additionally, interviews were conducted pertaining to the descriptive experiences of patients with depression and concerning the validity of adding items related to sensory awareness (E. Correa, personal interview, January 23, 1993; N. Pauker, telephone interview, February 12, 1993; S. Simpson, personal interview, February 26, 1993).

To analyze patterns and themes in depression, the next step involved a conceptual analysis of depression based on its DSM-III-R

definition and based on a comprehensive literature review. These themes were then clustered into three factors: physiological symptoms, emotional distress, and cognitive disturbance.

The subsequent step involved writing the items. The universe of content from which items could be written was identified, and, from a review of the literature, a content grid was developed. Together with Correa, this investigator then developed an initial 30-item pool (Appendix E).

Phase One: Qualitative Assessment

The next step of the study involved obtaining feedback from a panel of expert reviewers and from a focus group interview. Since the scale was intended to be used by an interdisciplinary team, the panel was comprised of different health care providers: three board certified psychiatrists, a doctorally prepared health educator, a psychometric psychologist, and a doctorally prepared psychiatric nurse. Decisions about scale revision were based upon the qualitative study findings.

Focus Group Interview

Basch defined the focus group interview as a "qualitative approach to learning about population subgroups with respect to conscious, semiconscious and unconscious psychological and sociocultural characteristics and processes" (1987, p. 411). Basch described it as a method used to obtain from small groups of

participants certain data about feelings and opinions concerning a given problem, experience, service or other phenomenon. Using nonprobability, purposive sampling, and thereby limiting generalizability of results, the focus group interview was determined to be an appropriate method for pilot-testing.

Basch further suggested several components for conducting a focus group interview: (a) using a group moderator to facilitate discussion from a prepared list of topics or questions; (b) utilizing the researcher as the instrument for promoting an atmosphere of physical comfort and trust thus encouraging freedom of expression; (c) tailoring subject recruitment to the goals of the research study; (d) limiting the focus group size to six to twelve participants; (e) having the interview last about one to three hours, and (e) utilizing a moderator to facilitate a free and open discussion about the concerns of the group.

For this research, focus group interviews were conducted on two different occasions, the first of which consisted of four university faculty, while the second was conducted with four registered nurses from a county health department. The purpose of the focus group was to obtain feedback not only about the clarity of the survey and its item construction but also about whether or not the survey's cover letter had sufficient appeal to motivate respondent completion. Results of the focus group interviews will be discussed in Chapter Four.

Expert Panel

Face and content validation of the Correa-Barrick Depression Scale and Depression Check-up was sought from an interdisciplinary panel (Appendix D) composed of three psychiatric experts, one academic psychologist with expertise in psychometrics, one health educator, and one psychiatric nurse.

To assess face and content validity, each expert panel member was sent a letter (Appendix F) requesting that the scales be reviewed to determine whether individual items measured depression. Content validity was assessed by asking the reviewers to determine whether the scales included criteria of the DSM-III-R. To maximize the response rate, follow-up letters were sent to each panel member (Appendix G). Results are discussed in Chapter Four.

Phase Two: Pilot Sample

Sample Population

The pilot sample consisted of a convenience sample of 100 students and 16 faculty and staff members from an academic department of a metropolitan comprehensive university.

The Depression Check-up and the Correa-Barrick Depression Scale were piloted on 100 undergraduate students. The purpose of the pilot was twofold:

1. Conduct a preliminary psychometric assessment of the Depression Check-up and the Correa-Barrick Depression Scale based on classical validity and reliability procedures and psychometric

comparisons with two published scales: the Beck Depression Inventory (Beck, Ward, Mendelson, Mock & Erbaugh, 1961) and the Inventory of Depressive Symptomatology, Self-report (IDS-SR; Rush et al., 1986).

2. Evaluate item-by-item reliability of the Correa-Barrick Depression Scale in order to edit or revise it to enhance scale reliability.

Pilot Procedure

Administration of the survey was conducted during class time in order to maintain security of the Beck Depression Inventory, a requirement of the test publisher (Appendix H).

On two different occasions, the Depression Check-up and the Correa-Barrick Depression Scale were piloted on 16 faculty and staff members from the university: (a) to pilot assessment of reliability and validity of the Depression Check-up and the Correa-Barrick Depression Scale and (b) to assess potential problems with the two-week test-retest administration procedure which would be conducted in the main study to evaluate scale stability over time.

Pilot Data Analysis

Pilot study outcomes were analyzed for item-scale correlations. As a result, troublesome items such as those with low and/or negative item correlations were revised or rejected in order to improve reliability of the Correa-Barrick Depression Scale.

To assess reliability, Cronbach's alpha and split-half (odd/even) correlation coefficients were computed. To measure convergent validity, correlation coefficients were produced for the Depression Check-up and the Correa-Barrick Depression Scale with the Beck Depressive Inventory and the IDS-SR. To evaluate concurrent validity, correlations between total depression scores for the Depression Check-up and the Correa-Barrick Depression Scale were computed with the scale items measuring severity of depression.

Phase Three: the Main Study

Sample Population

The main study was composed of two different groups. Sample I consisted of a convenience sample of all faculty and staff (n=1200) employed at a metropolitan comprehensive university and a subset sample of Sample I (hereafter called "Sample I Subset") which consisted of the respondents who returned the survey after the second mailing. Since the investigator was interested in determining reliability over time of the Depression Check-up and the Correa-Barrick Depression Scale and, in order to compute test-retest correlation coefficients between Sample I and the Sample I Subset, identical surveys were mailed at two different times (labeled "Time 1" and "Time 2").

The mailing list for Sample I originated from the university's office of human resources. Selected as the

"nonpatient" sample, this population was appropriate for the study because it represented a diversity of adult age groups and gender mix, as well as a variety of socioeconomic categories that included permanent full-time and part-time faculty, clerical staff, professionals, technicians (paraprofessional), skilled workers (electricians), graduate assistants, and student help. Cafeteria staff, housekeepers, and contractual personnel were not included.

Sample II consisted of 200 patients who had been diagnosed with major depression by a board certified psychiatrist according to the criteria for major depression as outlined in the DSM-III-R.

Main Study Procedure

The survey, consisting of background and demographic information, the Depression Check-Up, the Correa-Barrick Depression Scale, and the IDS-SR were distributed through campus mail on September 7, 1993 (Time 1), to the Sample I population (1,200 faculty and staff) with instructions for them to complete and return it on or before September 17, 1993 (Appendix B). The identical survey was mailed again to Sample I on September 22, 1993 (Time 2) with instructions for its completion and return on or before September 29, 1993. Respondents who returned the surveys at Time 2 comprised the Sample I Subset. A survey was also mailed to the investigator in order to trace the length of time it took for the surveys to be received (one working day).

Since the surveys were anonymous and time sensitive, sending follow-up letters to nonrespondents was not feasible. Therefore, the expected response rate was 30% for Sample I and 30% for the Sample I Subset from the total Sample I return. The minimum acceptable return needed for the validation study was 300 for the nonpatient population and 50 for the patient population. This acceptance rate was based upon the ratio of subjects-to-scale items required for factor analysis, for which the minimum number required was five to ten subjects per item, up to 300 subjects (DeVellis, 1991).

Test-retest of the survey was done on Sample I and the Sample I Subset to evaluate stability over time of the Depression Check-up and the Correa-Barrick Depression Scale. The nonpatient population was not expected to show changes in level of depression, since they were not experiencing treatment effects. Sample I was appropriate for measuring instrument stability, since a patient population may have shown changes in depression severity due to treatment effects which would confound the assessment of the instrument's reliability. Accordingly, a test-retest was not administered to Sample II.

To track surveys for the test-retest reliability procedure for Sample I and to maintain anonymity, each participant coded their own survey with the last two digits of their home telephone number followed by the last two digits of their social security number in a four-space coding on the first page of the survey. In

addition, each survey was color-coded: Time 1 surveys were orange, while Time 2 surveys were white.

To promote a good return rate, a cash prize incentive was awarded via mail in a random drawing once respondents had returned both surveys. A chart of matched participant codes was used for the purpose of verifying that a participant had returned both surveys.

The survey, consisting of background/demographic information, the Depression Check-up, the Correa-Barrick Depression Scale, and the IDS-SR, was also administered to 200 persons who, according to a psychiatrist's diagnosis, met the DSM-III-R criteria for depression. The surveys, along with a cover letter (Appendix I), were individually administered by the office manager as each patient arrived at the office for an appointment with the psychiatrist.

During initial administration of the surveys, the investigator spent three days in the waiting area to observe anonymously the administration process and respondent reactions. The office manager verbally instructed each patient that completion of the survey was voluntary and that the patient could stop at any time. Every effort was made to avoid taxing patients. When each respondent completed the survey, individual replies were viewed by the attending psychiatrist and not the investigator. At the time of their scheduled appointment, patients were individually debriefed by the psychiatrist about their reactions to the survey.

Collection of data from 50 patients took several months. Once data collection was completed, the psychiatrist channelled the data to the investigator for analysis.

Main Study Data Analysis

For the main study, data for the Depression Check-up and the Correa-Barrick Depression Scale were analyzed according to validity and reliability testing as described in Crocker and Algina (1986). The Statistical Package for the Social Sciences (SPSS) was used for data analysis procedures. Item analysis was executed for Sample I and Sample II. Because a full sample consisting of samples I and II would have the most range in scores, Samples I and II were merged, and item analysis was again repeated for the combined sample.

There were two procedures used for assessing construct validity; the first procedure assessed convergent and divergent validity, and the second used factor analysis.

To assess convergent validity, the Depression Check-up and Correa-Barrick Depression Scale were correlated with the BDI for the pilot sample only. Another correlation was conducted using the IDS-SR in Samples I and II (main study). Convergentvalidity of the Depression Check-up and the Correa-Barrick Depression Scale was assessed in the main study using t-tests to determine whether the item scores and total scores were significantly different for the patient population versus the nonpatient population. Discriminant

function analysis was also used to determine whether the two scales discriminated between the patient and nonpatient samples.

Factor analysis was performed on Samples I and II for the main study to determine whether there were three or four dimensions: cognitive disturbance, physiological symptoms, emotional distress, and sensory-perceptual disturbances for the Depression Check-up and the Correa-Barrick Depression Scale. Principal axis factoring method used varimax rotation. The factor analysis was repeated, allowing SPSS to determine the number of factors. Eigenvalues greater than 1.00 were used as the criteria for inclusion.

Several reliability analyses were performed on the Correa-Barrick Depression Scale and the Depression Check-up by the SPSS. (Interrater reliability was not assessed since this was not applicable for the self-report scale.) To assess internal consistency, Cronbach's alpha, which is appropriate for non-dichotomous variables, was assessed on Sample I and Sample II. To determine stability of the scales from a single-test administration, split-half (odd-even) was used. Spearman-Brown, which assumed that the half-tests were strictly parallel, was computed. The Depression Check-up contained 22 items, and the Correa-Barrick Depression Scale, 30 items. It was determined that both scales were of appropriate length for this procedure.

To assess stability over two weeks and to analyze stability coefficients using Pearson product moment correlation coefficients,

the Depression Check-Up and the Correa-Barrick Depression Scale were administered again at Time 2 to the Sample I Subset.

Since the Depression Check-up and Correa-Barrick Depression Scale were developed for screening purposes, preliminary norms based upon the samples used in this study were conducted to establish cut-off scores.

Summary

In this chapter, the methodology for validating the Depression Check-up and the Correa-Barrick Depression Scale, using both qualitative and quantitative assessment methods, was outlined. The next chapter will describe the results.

CHAPTER IV

DATA ANALYSIS, FINDINGS, AND INTERPRETATION

Introduction

The purpose of this study was twofold: to develop the Correa-Barrick Depression Scale and to establish initial validity and reliability of two new self-rating depression scales, the Depression Check-up (Schiraldi, 1987; Appendix A) and the Correa-Barrick Depression Scale (Appendix B) on two different population samples using survey research design.

Prior to this study, the Correa-Barrick Depression Scale was a new instrument and the Depression Check-up was in final form. Therefore, the purpose of this chapter is to discuss the three phases of development and refinement of the Correa-Barrick Depression Scale. Phase I involved qualitative analysis on the Correa-Barrick Depression Scale based upon focus group interviews and expert panel reviews. Phase II involved scale refinement for the Correa-Barrick Depression Scale and initial scale validity and reliability statistics for both the Correa-Barrick Depression Scale and the Depression Check-up based upon the pilot study findings. Phase III involved psychometric assessment of the Depression

118

Check-up and the Correa-Barrick Depression Scale and further scale revisions for the latter based upon the main study findings.

Research questions for the pilot study were:

1. Will the Depression Check-up and the Correa-Barrick Depression Scale possess convergent validity with the Beck Depression Inventory?

2. Will the Depression Check-up and the Correa-Barrick Depression Scale possess convergent validity with the Inventory of Depressive Symptomatology?

3. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate concurrent validity?

4. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability with internal consistency?

5. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability using the coefficient of stability from a single-test administration?

There were six research questions for the main study:

1. Will the Depression Check-up and the Correa-Barrick Depression Scale possess convergent validity?

2. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate concurrent validity?

3. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability based upon internal consistency?

4. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability over time?

5. Will the Depression Check-up and the Correa-Barrick Scale demonstrate construct validity?

6. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate discrimination?

Data analysis and findings for each of the research questions will be presented in the next sections. Prior to analysis, the database file was checked for data entry errors by conducting range checks. Two data entry errors were found and corrected.

Phase One: Qualitative Study Findings

The purpose of Phase I of the study was to conduct focus group interviews and expert panel reactions to the initial Correa-Barrick Depression Scale.

Focus Group

For this research, focus group interviews were conducted on two different occasions, the first of which consisted of four university faculty, while the second was conducted with four registered nurses from a county health department. The purpose of the focus group was to obtain feedback not only about the clarity of the survey and its item construction but also about whether or

not the survey's cover letter had sufficient appeal to motivate respondent completion.

Four faculty at Towson State University and four staff nurses at Eastern Family Resource Center, a division of Baltimore County Health Department, were interviewed after completing the research survey which included the Depression check-up (DC), the Correa-Barrick Depression Scale (CBDS), and the Inventory Depression Symptomatology (IDS).

Participants were asked to respond to the following question: "Overall, what was your reaction to the survey?" Based upon the investigator's anecdotal notes, respondents felt that the survey directions were clear, and all scales were short and easy to complete. Some respondents favored the Depression Check-up and the Inventory for Depressive Symptomatology, Self-report scales (Rush et al., 1986), since they provided very specific and descriptive item anchors. A few respondents reported difficultly discerning what number they should circle on the visual analogue scale for the Correa-Barrick Depression Scale, while others reported difficultly with the item on sexual activities and felt that the questionnaire did not take celibacy into consideration. One person noted that the item, "I still make decisions without any more difficulty than I usually have" was a double negative.

Individuals were asked about their perspective on providing a cash incentive to boost the survey return rate, and group members felt that an incentive plan was a good idea. Individuals reported

the use of colored paper would be appealing but reported not liking the idea of recording the last four digits of their home telephone numbers, since they believed that the true telephone number and identity of the respondent might be ascertained.

Expert Panel

An interdisciplinary panel composed of experts and academicians was sought to review the scale. Each panel member was mailed a packet which included draft copies of the Depression Check-up and the Correa-Barrick Depression Scale along with "Guidelines for Expert Reviewers" (Appendix K). When responses were not received by the due date, a telephone follow-up was conducted and a simplified cover letter was sent. The letter was designed so that responses could be written directly on it. This method was effective in producing written replies from three of the six panel members.

Dr. Ray DePaulo reviewed the Depression Check-up and found it to be "a reasonable set of items generally for a depression scale" (personal communication, February 12, 1993). However, he suggested splitting Item A ("Sleep disturbance [trouble getting to sleep or staying asleep, early wakening, or sleeping longer than usual]") and Item D ("A change in appetite, weight, or the amount you eat [a decrease or increase that you did not plan]"), since full weight was also given to items on "constipation," "headache," and "muscle tension." As an option, he recommended dropping these last three

on Item J ("Uninterested in sex"). He also noted that the Emotional subscale was probably not the best category, since Item E ("Unable to concentrate, think clearly, remember, or make decisions") was a thinking item.

Dr. DePaulo did not respond to the Correa-Barrick Depression Scale until six months later (personal communication, August 24, 1993). He validated both the face and content validity and agreed that the sensory items (color perception and taste) should be tested. He reported that the Correa-Barrick Depression Scale could be used for assessment of patient responses to pharmacological/psychological therapies and self-assessment, and he further concluded that it was "much improved" from the initial version.

For the Depression Check-up, Dr. Sylvia Simpson, another expert panel member, suggested that the term "working" on Item C under bodily symptoms (hyperactivity) be deleted and the terms "pacing" and "fidgety" be added (personal communication, February 26, 1993). She reported that the term "grouchy" should be added to Item G under emotional symptoms (Irritable: touch, nervous, jittery). She also thought that item E (unable to concentrate, think clearly, remember, or make decisions) under emotional symptoms should be moved under thought symptoms, since it is a cognitive item. Dr. Simpson determined that the Depression Check-up measured depression.

Dr. Simpson further concluded that the Correa-Barrick Depression Scale was a measure of clinical depression and reported that she was not aware of any literature findings on sensory alterations in depression. She did relate that sensory changes had been reported for mania and recommended that the sensory items be included in the Correa-Barrick Depression Scale. However, Dr. Simpson concluded that the items on "changes in sense of smell" should be deleted and felt that the psychomotor retardation item, "I have trouble talking because it takes so much effort," should be expanded to include additional terms.

In a telephone interview, Dr. Neil Pauker, one of the panel of experts, approved both the face and content validity of the Depression Check-up (personal communication, February 12, 1993). For the Correa-Barrick Depression Scale, Dr. Pauker suggested that the item on color perception ("I notice that everything seems gray/cloudy/drab/lacking color") be evaluated psychometrically and specifically suggested using the term "cloudy" on any color perception item. He also suggested that an item on "how food tastes" should be added, since many depressed patients do not eat because food does not taste the same. Dr. Pauker documented approval for both the face and content validity of the Correa-Barrick Depression Scale (personal communication, August 2, 1993).

Another expert panel member, Dr. Dianne Taylor, approved the face and content validity of the Correa-Barrick Depression Scale (personal communication, August 2, 1993). She noted that item #13

(I notice that food does not taste as good as usual) and item #14 (My sense of sound: music/conversation/nature/ is as sharp/clear as ever) appeared to be a part of depression but was not sure that these items were included on other depression scales (for example, the Beck Depression Inventory). Dr. Taylor also questioned whether item #17 (My ability to enjoy sex is worse than usual) meant that the ability to enjoy sex had never been good and was now worse. She also suggested adding hypersomnia to item #5 (I have sleep problems: trouble falling asleep/interrupted sleep/awaken too early).

Dr. Deitra Wengert, a member of the panel of experts, verbally approved the Depression Check-up scale and the Correa-Barrick Depression Scale "as is" (personal communication, September 4, 1993). Dr. John McGovern did not reply.

Use of Phase One Results

Based upon Phase I results (qualitative data) and Phase II pilot study results (quantitative data), the Correa-Barrick Depression Scale was revised for Phase III, the main study. Revisions are discussed under the section on "Scale Refinement."

Phase Two: Pilot Study

The purpose of Phase Two was twofold: to determine potential problems in implementing the main study and to use the results to refine the Correa-Barrick Depression Scale.

Sample characteristics

Surveys were administered to 100 students and 16 faculty at a comprehensive metropolitan university. Demographic data and sample characteristics from the pilot sample were computed from the SPSS-PC "cross tabulation" program as shown Table 1.

Table 1

Demographic and Selected Characteristics of Phase II:

Pilot Study Sample (n = 116)

Variable	Percent of Sample
Age	
Range 20 - 69 years	
Mean 35.2 years	
Gender	
Female	98.0
Male	2.0
Marital Status	
Single (never married)	43.1
Single (divorced)	6.9
Married, never divorced	37.1
Married, prior marriage(s)	6.9
Cohabitation	4.4
Widowed	1.7

(table continues)

Variable	Percent of Sample
Ethnic Group	
African American	12.1
American Indian	0.0
Asian	2.6
Hispanic	.9
Caucasian	84.5
Highest Level of Education	
Less than high school	0.0
High school	2.6
Some college	29.3
Associate Degree	9.5
Baccalaureate degree	44.0
Master's degree	5.2
Doctoral degree	7.8
Position	
Faculty	12.1
Professional staff	3.4
Staff	9.5
Student	73.3
Staff/Student	1.7
Family history of depression	
Yes	44.3
No	47.8
Don't Know	7.9

Variable	Percent of Sample
Personal history of depression	
Yes	18.1
No	81.9
Currently being evaluated for depression	
Yes	5.2
No	94.8
Currently on medications for depression	
Yes	3.4
No	96.6
Are you currently depressed	
Yes	6.9
No	81.0
Not sure	12.1

Preliminary Analysis

Before analyzing the data to answer the research questions, descriptive statistics for the Depression Check-up, the Correa-Barrick Depression Scale, the Beck Depression Inventory, and the Inventory for Depressive Symptomatology were produced and are summarized in Table 2.

Table 2

Descriptive Statistics: Pilot Sample

Scale	score values	mean score	median	SD	variance	S	n/obs	n/mv	SEM
Depression Check-up n = 116									
	1- 4	1.68	1.950	.4312	.1859	.9655	116	0	.0400
Correa-Barrick Depression Scale n = 116									
	0-10	3.75	4.480	.8562	.7332	.5096	116	0	.0795
Beck Depression Inventory n = 98									
	0- 3	.2757	1.285	.2781	.0774	1.7150	116	18*	.0280
Inventory for Depressive Symptomatology n = 116									
	0- 3	.46125	1.407	.3399	.1156	.9730	116	2**	.0318

SD   = standard deviation   var = variance   S   = skewness
n/obs = no. observations   n/mv = no. missing values   SEM = standard error of mean

*   For reasons of test security and contractual agreement with the test publisher, only 98 BDI surveys were administered to students in the pilot sample survey.
**   Number of missing values was expected because four of the scale items required respondents to choose only one item each from two paired responses.

Item Analysis

The Statistical Package for the Social Sciences program reliability was executed for previewing the item-total correlations and reliability coefficients. An examination of item-total correlations on the Correa-Barrick Depression Scale indicated several troublesome items. There were no negative item correlations on either the Depression Check-up or the Correa-Barrick Depression Scale. It should be noted that items numbered 4, 9, 10, 15, 25, and 26 on the Correa-Barrick Depression Scale were reverse coded (i.e., 10 = 0, 9 = 1, 8 = 2, 7 = 3, 6 = 4, 5 = 5, 4 = 6, and so forth.).

Based on item-total correlations, the range of item-total correlation coefficients for the 22-item Depression Check-up was .36 to .72. (n = 111), and the range for the 30-item Correa-Barrick Depression Scale ranged from .10 to .71 (n = 107). According to Nunnally (1967), a correlation coefficient at .6 - .7 was considered high; .4 - .5, medium, and any item less than .4, low. The investigator defined troublesome items as any item with an item-total correlation coefficient less than .40. Based on this criterion, 2 of the 22 items on the Depression Check-up would be considered troublesome: (a) Item #1--Sleep disturbance (trouble getting to sleep or staying asleep, early wakening, or sleeping longer than usual) and (b) Item #6--Headaches, or other aches and pains. In addition, 6 of 30 items on the Correa-Barrick

Depression Scale were less than .40. These items will be discussed in the next section.

Correa-Barrick Depression Scale Refinement

In addition to the six items on the Correa-Barrick Depression Scale that yielded correlation coefficients less than .40, all items with correlations less than .50 were scrutinized, and these items were revised based upon two criteria: correlations less than .50 and qualitative data from the focus groups. Based on these criteria, the following items were rewritten, and the correlation coefficients are reported in parentheses:

Item #2. "I have strange (vivid/weird) dreams and disturbing nightmares" ($r = .41$) was reworded to "I have noticed that my dreams are stranger (disturbing/vivid/weird) or that I have more nightmares."

Item #4. "My appetite is as good as ever" ($r = .34$) was reworded to "My appetite has changed (either increased or decreased)."

Item #7. "I am eating more/have increased cravings for sweets, snacks/am gaining weight" ($r = .26$) was changed to "I have increased cravings for sweets and carbohydrates."

Item #9. "I feel rested/pretty good in the morning" ($r = .49$) was changed to "I have trouble getting out of bed and getting started in the morning."

Item #14. "My sense of sound (music/conversation/nature) is as sharp/clear as ever ($r = .38$) was not changed, as the investigator decided to retain the exact item wording and test on the patient sample to see if the coefficient would increase.

Item #16. "My ability to enjoy sex is worse than usual" ($r = .38$) was changed to "My interest in being sensual/physical/loving is worse than usual." This item was reworded in order to facilitate a response category from participants who were celibate.

Item #20 "I still make decisions without any more difficulty than I usually have ($r = .41$) was changed to "I make decisions with difficulty."

Item #25. "My mood improves when something good happens" ($r = .09$) was changed to "I get excited/happy when something good happens."

Item #26. "I sometimes wish that I were dead" ($r = .46$) was changed to "I often have morbid thoughts (think about death or dying." This item was revised to delete the word "dead," a strong term that respondents may perceive as socially undesirable.

The final Correa-Barrick Depression Scale was a 30-item scale consisting of two components, Part A and Part B. Part A contained the items with the highest item-total correlation coefficients, those at .50 and above. Part B contained items below .50. Parts A and B were formatted in this manner in order to examine the patterns of item-total correlations from tne two different samples: the university and the depressed patient.

Preliminary Validity and Reliability

This section will discuss findings from the preliminary data analysis for the pilot sample. Each hypothesis will be discussed separately for the Depression Check-up and the Correa-Barrick Depression Scale. Unless otherwise noted, all quantitative results are reported in the positive (+) value.

Research Question 1. Will the Depression Check-up and the Correa-Barrick Depression Scale possess convergent validity with the Beck Depression Inventory?

Hypothesis 1-A. The Depression Check-up will positively correlate with the Beck Depression Inventory.

Hypothesis 1-B. The Correa-Barrick Depression Scale will positively correlate with the Beck Depression Inventory.

Based upon Pearson product moment correlation coefficients, results for the Depression Check-up with the Beck Depression Inventory were as follows: $r = .75$; $t = 11.09$; $df = 79$; $p < .01$. Results for the Correa-Barrick Depression Scale with the Beck Depression Inventory were: $r = .71$; $t = 9.94$; $df = 97$; $p < .01$.

Based upon these results, the null hypothesis was rejected; therefore, research Hypotheses I-A and I-B were accepted.

Research Question 2. Will the Depression Check-up and the Correa-Barrick Depression Scale possess convergent validity with the Inventory for Depressive Symptomatology?

Hypothesis 2-A. The Depression Check-up will positively correlate with the Inventory for Depressive Symptomatology, Self-report.

Hypothesis 2-B. The Correa-Barrick Depression Scale will positively correlate with the Inventory for Depressive Symptomatology.

Based upon Pearson product moment correlation coefficients, results for the Depression Check-up and the Inventory for Depressive Symptomatology were as follows: $r = .78$; $t = 13.07$; $df = 112$; $p < .01$. For the Correa-Barrick Depression Scale, the correlation coefficient with the Inventory for Depressive Symptomatology was: $r = .72$; $t = 11.05$; $df = 112$; $p < .01$. The null hypotheses was rejected, and research Hypotheses 2-A and 2-B were accepted.

Research Question 3. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate concurrent validity?

Hypothesis 3-A. The total depression score from the Depression Check-up will correlate with the following scale variable from the eighth question of the Depression Check-up: "Sad (gloomy, discouraged, blue, numb, empty, or like you just don't care)."

Hypothesis 3-B. The total depression score from the Correa-Barrick Depression Scale will correlate with the following scale variable from Question 21 of the Correa-Barrick Depression Scale: "I feel depressed (sad, blue, and gloomy)."

Results were similar for both the Depression Check-up and the Correa-Barrick Depression Scale when the item for depression was examined for its correlation with the total score. For the Depression Check-up, the item-total correlation coefficient for Question 8 with the total score was $r = .67$. Findings indicated that the item-total correlation for Question 21 on the Correa-Barrick Depression Scale and total score was $r = .71$.

Research Question 4. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability with internal consistency?

Hypothesis 4-A. The Depression Check-up will demonstrate a positive Cronbach's alpha coefficient.

Hypothesis 4-B. The Correa-Barrick Depression Scale will demonstrate a positive Cronbach's alpha coefficient.

The Cronbach's alpha for the Depression Check-up was $r = .92$, and, for the Correa-Barrick Depression Scale, it was $r = .92$ which indicated a high level of reliability for internal consistency.

Research Question 5. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability using the coefficient of stability from a single-test administration?

Hypothesis 5-A. The Depression Check-up will demonstrate a positive correlation coefficient on the split-half procedure (odd/even).

Hypothesis 5-B. The Correa-Barrick Depression Scale will demonstrate a positive correlation on the split-half procedure (odd/even).

For the Depression Check-up, the split-half Pearson product moment correlation coefficient was $r = .87$. The Spearman Brown correction was $r = .93$.

The split-half Pearson correlation for the CBDS was $r = .63$. The Spearman Brown correction was $r = .77$.

Interpretation of Pilot Study Findings

The pilot study findings suggested that both the Depression Check-up and the Correa-Barrick Depression Scale showed robust reliability coefficients for internal consistency based upon Cronbach's alpha. The Depression Check-up showed consistent results for all the split-half procedures, while the split-half reliability for the Correa-Barrick Depression Scale dropped to $r = .63$, but increased to $r = .77$ with the Spearman Brown.

Less impressive, but acceptable, were the concurrent correlations of the Depression Check-up and the Correa-Barrick Depression Scale with the two published depression scales. The Depression Check-up performed modestly higher than did the Correa-Barrick Depression Scale on convergence with both the Beck Depression Inventory and the Inventory for Depressive Symptomatology, Self-report. The convergent correlation of $r = .78$

for the Depression Check-up with the Inventory for Depressive Symptomatology, Self-report was considered better than acceptable.

Concurrent validity of the Depression Check-up and the Correa-Barrick Depression Scale with the specific scale items asking about level of depression were acceptable but not as high as expected. One factor that may explain the lower than expected findings was that the pilot sample did not include a known depressed sample. Coefficients between the items on the scales asking about depression and the total depression score would be expected to increase in a sample including depressed patients since the magnitude of the scores would vary to a greater extent.

Phase III: The Main Study

The purpose of this section will be to discuss the results of the main study which was conducted on two different sample populations: Sample I (a non-patient university sample) and Sample II (a depressed patient sample).

Sample I consisted of 1,200 permanent employees (faculty and staff) at a mid-Atlantic comprehensive university. Data was collected from Sample I at two different time periods, labeled Time 1 and Time 2. The Time 2 sample comprised a subset of Sample I and was used for the sole purpose of assessing test-retest reliability. Of the 1200 individuals asked to participate, 337 responded at Time 1, representing a 28% return rate. At Time 2, 203 surveys were returned, representing a 17% return rate.

Sample II consisted of all 200 patients in a private psychiatric outpatient setting. Criteria for inclusion in the patient sample consisted of a prior and/or current diagnosis of major depression as operationally defined by the Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R) of the American Psychiatric Association (APA, 1987). All determinations of diagnosis were made by a psychiatrist who was board certified in psychiatry and neurology. Surveys were distributed to the patients by the office manager and results were reviewed by the psychiatrist and this investigator. Two hundred surveys were administered individually to patients who met the DSM III-R diagnostic criteria for at least one episode of major depression, as determined by the attending psychiatrist. Fifty surveys were returned, representing a response rate of 25%. Tables 3, 4, and 5 show the demographic and selected characteristics of the main study sample.

In order to achieve a wider range of scores for two of the statistical procedures, namely item analysis and discriminant function analyses, Sample I and Sample II were combined to produce a combined sample set.

Table 3

Demographic and Selected Characteristics of Phase III: Main Study,

University Sample at Time 1 (n = 337)

Variable	percent of sample
Age	
mean 45.7 years	
range 22 – 69 years	
Gender	
female	56.6
male	43.4
not reported (n = 12)	
Marital status	
single (never married)	16.4
single (divorced)	8.6
married (never divorced)	53.0
married, prior marriages	20.8
widowed	1.2
not reported (n = 1)	
Ethnic group	
African American	7.8
American Indian	.3
Asian	2.7
Hispanic	.6
Caucasian	86.9
other	1.8
not reported (n = 2)	

Variable	percent of sample

Level of education
    less than high school   .9
    high school   1.0
    some college   15.2
    associate degree   2.1
    baccalaureate degree   18.5
    master's degree   19.0
    doctoral degree   33.3
    not reported (n = 1)

Position
    faculty   35.9
    professional staff   33.2
    staff   30.8
    not reported (n = 3)

Family history of depression
    yes   36.8
    no   56.1
    don't know   7.1

Personal history of depression
    yes   17.5
    no   82.5

Currently being evaluated for depression
    yes   6.5
    no   93.5

Currently on medication to treat depression
    yes   5.3
    no   94.7

(table continues)

Variable	percent of sample

Currently depressed

yes	8.3
no	83.1
not sure	8.6

Table 4

Demographic and Selected Characteristics of Phase III: Main Study,

University Sample at Time 2 (n = 203)

Variable	percent of sample

Age

    mean 46.1 years

    range 24 - 69 years

Gender

female	57.3
male	42.7
not reported (n = 18)	

Marital status

single (never married)	16.2
single (divorced)	9.4
married (never divorced)	55.2
widowed	20.8
not reported (n = 1)	

Variable	percent of sample
Ethnic group	
African American	7.8
American Indian	0
Asian	2.5
Hispanic	1.1
Caucasian	89.2
Level of Education	
less than high school	.4
high school	9.4
some college	15.8
associate degree	1.4
baccalaureate degree	19.8
master's degree	22.3
doctoral degree	30.9
Position	
faculty	32.1
professional staff	39.1
staff	30.8
not reported (n = 4)	
Family history of depression	
yes	35.3
no	59.0
don't know	5.8
Personal history of depression	
yes	18.3
no	81.7

(table continues)

Variable	percent of sample
Currently being evaluated for depression	
yes	5.8
no	94.2
Currently on medication to treat depression	
yes	5.0
no	95.0
Currently depressed	
yes	4.7
no	85.9
not sure	9.4
not reported (n = 1)	

Table 5

Demographic and Selected Characteristics of Phase III:

Patient Sample (n = 50)

Variable	Percent of sample
Age	
mean 45.9 years	
range 27 - 80 years	
Gender	
female	62.5
male	37.5
not reported (n = 2)	(table continues)

Variable	Percent of sample
Marital status	
single (never married)	22.0
single (divorced)	8.0
married (never divorced)	54.0
married (prior marriages)	8.0
widowed	8.0
Ethnic group	
African American	4.0
Caucasian	96.0
Level of Education	
less than high school	2.0
high school	10.0
some college	26.0
associate degree	4.0
baccalaureate degree	24.0
master's degree	28.0
doctoral degree	6.0
Family history of depression	
yes	64.0
no	24.0
not sure	12.0
Currently on medication to treat depression	
yes	96.0
no	4.0

(table continues)

Variable	Percent of sample
Currently depressed	
yes	32.7
no	44.9
not sure	22.4
not reported (n = 1)	

Patients who refused to participate reported to the psychiatrist that the survey was too lengthy; they were either too depressed to complete the survey or were disturbed by its length. To ascertain patient perceptions of the depression scales, patients were individually interviewed by the psychiatrist. The pattern of reported responses indicated that patients felt overwhelmed by the Inventory for Depressive Symptomatology scale (IDS-SR), because it was too wordy and too long. Specifically, the IDS scale is similar to a multiple-choice test format in that it requires the patient to decide from a four-item response category labeled A, B, C, D. The Correa-Barrick Depression Scale's visual format and the Depression Check-up were preferred because of design simplicity and deemphasis on vocabulary. Patients felt that the Depression Check-up and the Correa-Barrick Depression Scale were the least emotionally taxing to complete.

Interpretation of Findings: Survey Response

It was interesting to note that faculty and staff seemed to prefer the IDS, perhaps because its vocabulary was at a higher level; whereas, depressed patients reported a preference for simplistic depression scales. The IDS may be measuring vocabulary which is more appealing to normal subjects who are well educated, and this may have been the case with the university sample. On the other hand, even well-educated depressed subjects may be too cognitively impaired to complete lengthy questionnaires.

Scale Descriptive Statistics

Descriptive statistics to show the average and range of the total depression scores from each scale were produced for the Depression Check-up, the Correa-Barrick Depression Scale, and the IDS for the university and patient samples. The results are summarized in Table 6. Missing data did not occur often and, therefore, was ignored. The remaining values in the data set were calculated for the statistical analysis except for item analysis. For item analysis, SPSS uses "likewise deletion of cases" on any missing values. Likewise deletion drops missing values from any of the scale items, thereby reducing the total number of cases used in the analysis. Thus, "n" varied on item analysis. The handling of missing values using likewise deletion of cases will be explained further under the section on item analysis.

Table 6

Descriptive Statistics: University Sample (n=337) and Patient Sample (n=50)

Scale	mean per item	median per item	SD per item	var per item	SEM per item	total # items	actual score*	possible range of scores
Depression Check-up								
University sample	1.580	1.450	.491	.242	.0268	22	35	22-88
Patient sample	2.040	1.90	.624	.389	.0820	22	45	22-88
Correa-Barrick Depression Scale								
University sample	3.246	3.030	1.020	1.040	.0550	30	97	0-300
Patient sample	3.750	4.480	1.870	3.510	.2650	30	109	0-300
Inventory for Depressive Symptomatology								
University sample	.442	.344	.367	.134	.0200	30	12	0-84
Patient sample	.871	.803	.489	.239	.0690	30	24	0-84

SD   = standard deviation   var = variance   SEM = standard error of measure
*actual score = mean score per item x number of items

Item Analysis: University Sample

Item-total correlation coefficients were produced using the "reliability" program. Sample numbers were reported for each item, but the total number of cases used for item analysis was different from the sample number because of the SPSS procedure, "likewise deletion of cases." For item analysis, any scale item which showed a missing value was deleted from the case count. It was noted that, for the Correa-Barrick Depression Scale, there were 12 cases missing from the item analysis due to "likewise deletion of cases." A total of 12 cases from the total sample of 50 were deleted because some of the items had missing values. Item analysis is a method for assessing internal consistency and, as such, requires that there be no missing values for any items. The procedure, "likewise deletion of cases," is an appropriate method for handling missing values.

Missing values should be of concern to the integrity of a study. Therefore, the investigator visually scrutinized the surveys to determine the reason or reasons for the missing values and found that item #14 on the Correa-Barrick Depression Scale contained 10 missing values. Unfortunately, item #14 was at the top of the second page closely aligned near the stapled area of the page; it is likely that this obscured item #14. Therefore, it was assumed that wording did not deter respondent completion of this item.

For the Depression Check-up, item-total correlation coefficients for the university sample ranged from r = .33 to r = .75. Item #5, "Constipation," had a correlation coefficient of r = .33. For the patient sample, item-total correlation coefficients ranged from r = .43 to r = .83 (Table 7).

Table 7

Depression Check-up: Item-Total Correlation Coefficients in the University and Patient Samples

	university sample		patient sample	
Item	n	r	n	r
1 sleep disturbance	(336)	.58	(49)	.63
2 tired feeling	(335)	.72**	(49)	.66
3 slowed activity	(335)	.67	(49)	.69
4 a change in appetite	(336)	.52	(50)	.43
5 constipation	(336)	.33*	(50)	.54
6 headaches	(336)	.56	(50)	.65
7 muscle tension	(336)	.54	(50)	.67
8 sad	(336)	.74**	(50)	.77**
9 things aren't fun	(336)	.74**	(50)	.78**
10 down on yourself	(336)	.70**	(50)	.83**
11 guilty (bad person)	(336)	.63	(50)	.72**

	university sample		patient sample	
Item	n	r	n	r
12 unable to concentrate	(336)	.64	(50)	.72**
13 like crying	(336)	.66	(50)	.75**
14 irritable	(336)	.64	(50)	.57
15 uninterested in people	(333)	.70**	(49)	.75**
16 unattractive	(335)	.75**	(50)	.72**
17 uninterested in sex	(335)	.58	(49)	.56
18 wrong with you	(335)	.72**	(50)	.77**
19 thinking . . . world	(336)	.56	(50)	.56
20 thinking . . . suicide	(335)	.50	(50)	.66
21 negative aspects	(335)	.68	(50)	.58
22 worrying about health	(336)	.65	(50)	.67

** r greater than .70

note: 332 = number of cases for item analysis, university sample
      47 = number of cases for item analysis, patient sample

The item reliability analysis for the Correa-Barrick
Depression Scale revealed low to high correlations, ranging from
r = .25 to r = .78. Items numbered 9, 10, 15, 21, 25, and 26 were
reverse-coded. As shown in Table 8, item-total correlation
coefficients for the Correa-Barrick Depression Scale ranged from

$r = .26$ to $r = .75$; two of the 30 items (21 and 23) were less than $r = .3$ in the university sample

Table 8

Correa-Barrick Depression Scale: Item-Total Correlation

Coefficients in the University and Patient Samples

Item		university sample		patient sample	
		n	r	n	r
1	I have sleep	(335)	.51	(48)	.54
2	I feel tired	(333)	.60	(48)	.71**
3	My body feels heavy	(336)	.74**	(49)	.70**
4	I am excited	(336)	.40	(48)	.57
5	Trouble with activity	(336)	.62	(49)	.65
6	I feel restless	(335)	.45	(49)	.59
7	wrong with health	(334)	.51	(48)	.74
8	lacking color	(335)	.46	(49)	.77**
9	notice attractive	(335)	.25*	(47)	.56
10	look forward	(335)	.47	(49)	.75**
11	bored and worry	(335)	.62	(49)	.67
12	taste not as good	(335)	.49	(49)	.53
13	I feel depressed	(335)	.74**	(49)	.78**
14	concentration	(335)	.60	(43)	.76**

Item	university sample		patient sample	
	n	r	n	r
15 look forward	(335)	.58	(50)	.58
16 more faults	(335)	.60	(50)	.62
17 feel afraid	(335)	.55	(50)	.61
18 irritable	(336)	.69	(50)	.54
19 feel like crying	(335)	.67	(50)	.71**
20 feel responsible	(333)	.60	(50)	.78**
21 losing weight	(335)	.25*	(50)	.37*
22 dreams	(336)	.39*	(50)	.35*
23 sweet/CHO cravings	(335)	.27*	(50)	.55
24 trouble in a.m.	(335)	.57	(50)	.54
25 sounds	(336)	.31*	(50)	.60
26 excited/happy	(336)	.21*	(49)	.43
27 morbid thoughts	(334)	.58	(50)	.77**
28 sensual/loving	(327)	.58	(50)	.76**
29 decision difficulty	(336)	.51	(50)	.77**
30 appetite has changed	(336)	.51	(50)	.58

* r less than .40
** r greater than .70

note: 318 = number of cases for item analysis, university sample
      38 = number of cases for item analysis, patient sample

After examining the lowest item-total correlation coefficients in the university sample on the Correa-Barrick Depression Scale and "alphas if the items deleted" from the reliability program, it was decided to delete six items from the original scale and identify this 24-item version as the Correa-Barrick Depression Scale--Short Version. (Later in this chapter, the statistical analysis for the 24-item Correa-Barrick Depression Scale--Short Version will be presented and compared with the original version. The listing below shows the six items with the lowest item-total correlations and corresponding alphas:

Item	Item wording	r	Alpha if item deleted
9	I notice attractive men and women.	.25	.91
21	I notice that I am losing weight.	.25	.91
22	I have noticed that my dreams are stranger (disturbing/vivid/weird) or that I have more nightmares.	.39	.91
23	I have increased craving for sweets and CHOs	.27	.92
25	My appreciation of sounds (music/ conversation/nature) is as good as ever..	.31	.91
26	I get excited/happy when something good happens.	.21	.92

Table 9 shows the descriptive statistics for the Correa-Barrick Depression Scale--Short Version.

Table 9

Descriptive Statistics: Correa-Barrick Depression Scale--Short Version

Scale	mean score	median	SD	variance	S	n/obs	n/mv	SEM
Correa-Barrick Depression Scale-- Short Version	2.87	2.62	1.19	1.43	1.01	337	1	.065

SD   = standard deviation   var   = variance   S   = skewness
n/obs = no. observations   n/mv = no. missing values   SEM = standard error of mean

Item analysis was repeated for the total sample which
included Sample I (university) and Sample II (patient). Item-total
correlation coefficients are presented in Table 10 for the
Depression Check-up and in Table 11 for the Correa-Barrick
Depression Scale.

Table 10

Depression Check-up: Item-total Correlation
Coefficients in Combined Sample (n = 379)

Item		r
1	Sleep disturbance	.60
2	Tired Feeling	.72**
3	Slowed activity	.70**
4	Change in appetite	.54
5	Constipation	.42*
6	Headaches	.59
7	Muscle tension	.57
8	Sad	.76**
9	That Things aren't fun	.77**
10	Down on yourself	.75**
11	Guilty (bad person)	.66
12	Unable to concentrate	.69
13	Like crying	.69

(table continues)

Item		r
14	Irritable	.64
15	Uninterested in people	.73**
16	Unattractive	.74
17	Uninterested in sex	.54
18	Thinking . . . wrong with you	.75**
19	Thinking world is harsh	.54
20	Suicide	.55
21	Noticing negative aspects	.67
22	Worrying about health	.65

Cronbach's alpha = .95

** *r* greater than .70

Table 11

Correa-Barrick Depression Scale: Item-total Correlation

Coefficients in Combined Sample (n = 356)

Item		r
1	I have sleep problems	.47
2	I feel tired	.63
3	My body feels heavy	.72**
4	I am excited	.39*
5	Trouble with activity	.66
6	I feel restless	.52
7	Wrong with health	.59
8	Lacking color	.55
9	Notice attractive	.32*
10	Look forward	.45
11	Brood and worry alot	.62
12	Taste not as good	.52
13	I feel depressed	.75**
14	Concentration	.63
15	Look forward	.55
16	More faults	.63
17	Feel afraid	.59
18	Irritable	.66

(table continues)

Item		r
19	Feel like crying	.68
20	Feel responsible for bad things	.62
21	Losing weight	.28*
22	Dreams strange	.45
23	Sweet/CHO cravings	.31*
24	Trouble in a.m.	.60
25	Sounds	.37*
26	Excited/happy	.26*
27	Morbid thoughts	.59
28	Sensual/physical/loving	.61
29	Decision difficulty	.56
30	Appetite has changed	.54

Cronbach's alpha = .92

*   r less than .40

**   r greater than .70

Interpretation of Findings: Item Analysis

The pattern of item-total correlations showed higher item-total correlations in the patient sample than in the university sample. However, appetite change and weight loss items were low in

both samples. Items related to emotional variables such as "sad and depressed" and cognitive variables such as "things aren't fun anymore" and self-esteem items showed high item-total correlation in both scales. Color perception alteration, $r = .77$, was high in the patient sample for the Correa-Barrick Depression Scale thus suggesting an additional variable in depression.

The item-total correlation coefficient patterns remained essentially unchanged in the total sample. Items #9, #21, #22, #23, #25, and #26, which were low in the university patient sample, were also consistently low in the combined sample with one exception, item #23. It was noteworthy that item #23, "Cravings for sweets and CHO," had an item-total correlation $r = .27$ in the university sample, $r = .31$ in the combined sample, and $r = .55$ in the patient sample. This finding suggested that cravings for sweets and CHO may be an additional variable reflecting an atypical symptom in depression and/or may have been a happenstance finding.

In addition to comparing the quantitative item analysis results for the Depression Check-up and the Correa-Barrick Depression Scale, a conceptual analysis of the variables on the scales may also provide useful foundational information for later factor analyses. Accordingly, a qualitative analysis of all item-total correlation coefficients (r) for both the Depression Check-up and Correa-Barrick Depression Scale in the combined sample set was scrutinized for similar values. Items that yielded high value r for the Depression Check-up and the Correa-Barrick

Depression Scale were examined to determine whether the item wording was comparable in content. Once comparable items from each scale which matched in similar r values were found, the items were clustered into different conceptual areas. This analysis suggested several conceptual areas: energy, sad and depressed mood, self-esteem, crying, excessive guilt, concentration, irritability, and outlook.

The same analysis was repeated for the lowest item-total correlations for the Depression Check-up and Correa-Barrick Depression Scale. It was found that the item-total correlations for biological symptoms such as weight loss, appetite, and sleep items were not as high as expected.

Validity and Reliability

This section will discuss findings from data analysis for the main study. Each hypothesis will be discussed separately for the Depression Check-up and the Correa-Barrick Depression Scale.

Research Question 1

Will the Depression Check-up and the Correa-Barrick Depression Scale possess convergent validity with a known published scale?

Hypothesis 1-A. The Depression Check-up will positively correlate with the Inventory for Depressive Symptomatology, Self-report. (Supported)

The Pearson correlation coefficient for the Depression Check-up with the Inventory for Depressive Symptomatology was $r = .88$, $t = 34.31$, $df = 333$, $p < .01$ for the university sample. The Pearson correlation coefficient for the patient sample was $r = .85$, $t = 11.30$, $df = 48$, $p < .01$.

Hypothesis 1-B. The Correa-Barrick Depression Scale will positively correlate with the Inventory for Depressive Symptomatology, Self-report. (Supported)

For the university sample, the correlation coefficient for the Correa-Barrick Depression Scale with the Inventory for Depressive Symptomatology was $r = .72$, $t = 18.71$, $df = 333$, $p < .01$.

This analysis was repeated with six of the 30 Correa-Barrick Depression Scale items deleted, as listed previously in this chapter, to ascertain whether the coefficient with the Inventory for Depressive Symptomatology would improve. When these six items were deleted, there was only modest improvement. The correlation between the Correa-Barrick Depression Scale and the Inventory for Depressive Symptomatology increased from $r = .72$ to $r = .77$, $p < .01$.

For the patient sample, the Pearson correlation coefficient between the 30-item Correa-Barrick Depression Scale and the

Inventory for Depressive Symptomatology improved in the patient sample from $r = .72$ to $r = .81$, $t = 9.55$, $df = 48$, $p < .01$.

For the patient sample, the Pearson correlation coefficient between the 30-item Correa-Barrick Depression Scale and the Inventory for Depressive Symptomatology, Self-Report improved in the patient sample from $r = .72$ to $r = .81$, $t = 9.55$, $df = 48$, $p < .01$. Correlation coefficients for the university and patient samples are summarized in Tables 12 and 13.

Table 12

Correlation Matrix: Comparisons in University Sample and Patient Sample (n = 50)

			IDS-SR	
	CBDS	DC	university n=337	patient n=50
CBDS	1.00	.77	.72*	.81*
DC	.77	1.00	.88*	.85*
IDS	.72	.88	1.00	1.00

*   p < .01

CBDS   Correa-Barrick Depression Scale
DC   Depression Check-up
IDS-SR   Inventory for Depressive Symptomatology, Self-report

Table 13

Comparison of Convergent Validity Coefficients for the DC and CBDS

with IDS in University Sample and Patient Sample

	r (shared variance)	
	university sample	patient sample
DC	.88* (78%)	.85 n.s. (71%)
CBDS	.72* (52%)	.81 n.s. (66%)

* $p < .01$

n.s.   not significant

Tests for assessing whether there were significant differences between the Depression Check-up and the Correa-Barrick Depression Scale correlation coefficients in the university sample were calculated using the Fisher z transformation test. Results showed $z = 6.45$, $p < .01$.

The z test for the patient sample was .625, which was not significant, indicating there were no differences between correlation coefficients when comparing the Depression Check-up and the Correa-Barrick Depression Scale.

Interpretation of Findings

Based on the findings, there was evidence to support Research Question 1. It should be noted that the Depression Check-up correlation coefficients with the Inventory for Depressive Symptomatology were higher than that of the Correa-Barrick Depression Scale with the IDS in the university sample. The Depression Check-up also demonstrated more stability (similar findings) than the Correa-Barrick Depression Scale across both the university sample and the patient sample. Specifically, the convergent validity of the Depression Check-up with the IDS was similar in both the university and patient samples. However, this was not the case with the Correa-Barrick Depression Scale, since the correlation coefficient with the IDS was much lower in the university sample than in the patient sample.

One possible explanation for the inconsistent findings in the different samples for the Correa-Barrick Depression Scale may have been due to the scale's visual analogue format which was not driven by verbal anchors to the extent that the Depression Check-up and, to a greater extent, the IDS was. Verbal anchors may viewed as more accurate reflections of feeling states in a normal, well-educated sample than in a psychiatrically depressed sample. In contrast, depressed patients may respond more consistently to visual anchors.

Lastly, the findings that emerged when comparing the Depression Check-up and Correa-Barrick Depression Scale in the

university sample was unexpected. Specifically, significant differences in the correlation coefficients were not expected. However, a visual examination of the correlation coefficients seemed to suggest practical differences as well.

Additional Analysis

Further analysis was performed to determine whether there were group differences, based on marital status, gender, and age, on the total depression scores for the Depression Check-up and the Correa-Barrick Depression Scale. As reported in Chapter II, the literature consistently showed, on replicated cross-cultural studies, that age, marital status, and gender had a differential effect on depression rates. To determine whether there were group differences in the Depression Check-up and Correa-Barrick Depression Scale, marital status, age, and gender were scrutinized for differences in the university and patient samples.

For the Depression Check-up and the Correa-Barrick Depression Scale, two sample t-tests were computed on the university sample to determine whether there were significant group differences for marital status and gender. Several adjustments were made to the data for the analysis. For the university sample (n = 335), the cell size for marital status (widowed) was small (n = 4); therefore, this category was filtered out for the analysis. For the patient sample (n = 50), only the two largest cells in the marital status category (single, never married and married, never

<u>divorced</u>) were used, since the other cell sizes (<u>divorced</u>, <u>prior</u> <u>marriages</u>, and <u>widowed</u>) were small for each cell (n = 4).

Results indicated there were no significant group differences on marital status and gender for the Depression Check-up and for the Correa-Barrick Depression Scale in both the university and patient samples.

Since the variable <u>age</u> and the Depression Check-up and Correa-Barrick Depression Scale scores were interval-level data, multiple regression, using age as the dependent variable, was computed. There was no significant difference for the Depression Check-up on age; however, there was a significant difference for the Correa-Barrick Depression Scale for age, as shown in Table 14.

Table 14

Results of Multiple Regression on CBDS, Age as Dependent Variable

Source	df	SS	MS	F-ratio	
model	1	452.0058	462.0058	4.75	p = .029*
error	311	30252.4000	97.27461		
total	312	30714.4100	98.44362		

r	= .1225
R squared	= .0150
adjusted R squared	= .0119

* p < .05

Although the Correa-Barrick Depression Scale demonstrated an effect of age on total depression score, the variance (expressed as R squared), albeit statistically significant, seemed to be practically insignificant. Group differences failed to emerge for marital status, gender, and age for the Correa-Barrick Depression Scale and for the Depression Check-up in the university and patient samples.

Research Question 2

Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate concurrent validity with the variables that are expected to correlate with depression?

Hypothesis 2-A. The total depression score from the Depression Check-up will positively correlate with the following survey variables: family history, personal depression history, taking medications for depression, current depression, and question II-A of the Depression Check-up: "Sad (gloomy, discouraged, blue, numb, empty, or like you just don't care)." (Supported for family history, currently depressed, and question II-A.)

Hypothesis 2-B. The total depression score from the Correa-Barrick Depression Scale will positively correlate with the following survey variables: family history, personal depression history, taking medication for depression, and Question 13 of the Correa-Barrick Depression Scale: "I feel depressed (sad, blue, and

gloomy)." (Supported only for family history, currently depressed, and Question 13.)

For both scales, ANOVAs were performed for all the survey variables. There were no significant group differences for personal depression history ("Have you ever been evaluated, referred, or treated for depression?") or for depression medication ("Taking medications to treat depression"). Results showed there were significant group differences on "Family history of depression" and "Are you currently depressed" for both the Correa-Barrick Depression Scale (Table 15) and the Depression Check-up (Table 16).

Table 15

Analysis of Variance: Correa-Barrick Depression Scale, University Sample (n = 336)

Source of variance	df	SS	MS	F ratio	probability
family depression	2	19.07	9.53	9.60	.0001*
error	333	331.00	.99		
Total	335				

*p < .01

Table 16

Analysis of Variance: Depression Check-up,

University Sample (n = 336)

Source of variance	df	SS	MS	F ratio	probability
family depression	2	6.29	3.14	14.01	.0000*
error	333	74.70	.22		
Total	335	81.07			

*p < .01

Table 17

Analysis of Variance for Correa-Barrick Depression Scale (n = 336)

Source of variance	df	SS	MS	F ratio	probability
currently depressed	2	74.5	37.20	45.08	.0000*
error	333	275.5	.83		
Total	335	350.1			

*p < .01

Table 18

Analysis of Variance: Depression Check-up (n = 336)

Source of variance	df	SS	MS	F ratio	probability
currently depressed	2	28.60	14.30	90.88	.0000*
error	333	52.46	.15		
Total	335	81.07			

*p < .01

Results for question #13 on the Correa-Barrick Depression Scale ("I feel depressed") showed a correlation coefficient, r = .68, p < .01 with the total score for the university sample and a correlation coefficient, r = .83, p < .01 for the patient sample. Results for the Depression Check-up question about feeling "Sad (gloomy, discouraged, blue, numb, empty, or like you just don't care)" showed r = .75, p < .01 for the university sample and r = .78, p < .01 for the patient sample.

Interpretation of Findings

The findings inconsistently supported Research Question 2 for concurrent validity with the survey variables that were expected to correlate with depression: "family history," "personal depression history," "taking medication for depression," "are you currently depressed," and the Depression Check-up and the Correa-Barrick

Depression Scale items assessing level of depression. Of the variables, only family history and current depression showed significant group differences. Nonetheless, the Depression Check-up and Correa-Barrick Depression Scale items which assessed level of depression showed signficant correlation with total depression score. These findings would tend to partially support concurrent validity of the Depression Check-up and the Correa-Barrick Depression Scale.

Research Question 3

Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability based upon internal consistency?

Hypothesis 3-A. The Depression Check-up will demonstrate a positive Cronbach's alpha. (Supported)

Hypothesis 3-B. The Correa-Barrick Depression Scale will demonstrate a positive Cronbach's alpha. (Supported)

The first step involved item analysis for each scale by computing item-total correlation coefficients for both scales. This was accomplished and reported under preliminary analysis.

For reliability analysis, the average of all split-halves was examined using Cronbach's alpha. The Statistical Package for the Social Sciences (SPSS) treated missing values with likewise deletion of cases.

For the university sample (n = 332), Cronbach's alpha for the Depression Check-up was r = .94. For the university sample (n = 318), Cronbach's alpha for the Correa-Barrick Depression Scale was r = .93. For the patient sample (n = 38), Cronbach's alpha for the Correa-Barrick Depression Scale was r = .96. For the patient sample (n = 47), Cronbach's alpha for the DC was r = .95.

When the six low items showing a low item-total correlation on the Correa-Barrick Depression Scale were deleted, Cronbach's alpha remained relatively unchanged: alpha coefficient for the university sample (n = 318) and for the patient sample (n = 38) was r = .95.

Interpretation of Findings

The results supported Research Question 3 and suggested internal consistency for both the Depression Check-up and the Correa-Barrick Depression Scale. To verify the findings, the analysis was repeated for both the Depression Check-up and for the Correa-Barrick Depression Scale with the same results.

Research Question 4

Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability over time?

Hypothesis 4-A. Depression scores on the Depression Check-up will positively correlate with the depression scores on

the Depression Check-up when administered two weeks or 14 days later. (Supported)

Hypothesis 4-B. Depression scores on the Correa-Barrick Depression Scale will positively correlate with the depression scores on the Depression Check-up when administered two weeks or 14 days later. (Supported)

To ascertain stability of the Depression Check-up and the Correa-Barrick Depression Scale over a two-week time period, Pearson product moment correlation coefficients were used to compute correlations between the university sample at "Time 1" and the university sample two weeks later ("Time 2") for matched pairs (n = 203).

Correlation coefficient results for the Depression Check-up were $r = .81$, $p < .01$. Results for the Correa-Barrick Depression Scale were $r = .70$, $p < .01$.

The procedure was repeated for the Correa-Barrick Depression Scale, removing troublesome items from the analysis to determine whether results would change if such items were removed from the Correa-Barrick Depression Scale. Results for the repeat analysis showed that the test-retest correlation coefficients improved modestly from $r = .70$ to $r = .73$, p. $< .01$.

Results for the published scale, the Inventory for Depressive Symptomatology, Self-Report (IDS-SR), showed a test-retest correlation coefficient $r = .71$, $p < .01$.

Interpretation of Findings

Based on the results of the DC and the CBDS and the comparison with the published IDS, the data support the hypotheses that both scales demonstrate stability over time in a university sample (N= 203). The Depression Check-up, yields better stability over time. The CBDS shows a lower stability over time which may be due to either lower stability over time or the visual analog format. Specifically, as reported in chapter two, the Visual analogue scale (VAS) tends to impair respondent recall of prior answers (Gift, 1989).

Research Question 5

Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate construct validity?

Hypothesis 5-A. The Depression Check-up will cluster on three dimensions from factor analyses: cognitive impairment, physiological symptoms, and emotional distress. (Supported only for a 3-factor structure.)

Hypothesis 5-B. The Correa-Barrick Depression Scale will cluster on four dimensions from factor analyses: cognitive impairment, physiological symptoms, emotional distress, and impaired sensory awareness. (Supported only for a 4-factor structure.)

Factor Analysis for the Depression Check-up

The first step in the factor analysis involved producing initial, unrotated statistics by the method of principal components analyses. This method extracted three factors with eigenvalues greater than 1.0, which was consistent with theoretical expectations, resulting in a cumulative variance of 57% (Table 19).

Factor loadings for Factor 1 ranged from .86 to .79; Factor 2 ranged from -.86 to .48, and Factor 3 ranged from -.38 to .62. The initial statistics and eigenvalue patterns are in shown in Table 19.

The second step was to decide which type of rotation to use. It was decided to use an orthogonal rotation using the Varimax, since the investigator assumed that the factors are not correlated. This procedure also resulted in the simplest factor matrix to interpret. Varimax rotation resulted in 7 iterations with a rotated factor matrix comprised of three (3) factors. Factor 1 comprised loadings ranging from .12 to .85. Factor 2 had loadings ranging from -.22 to .73. Factor 3 comprised loadings ranging from .01 to .73 (Table 20).

Table 19

Depression Check-up: Initial Statistics and Eigenvalue Summary,

University Sample (n = 337)

variable	factor	eigenvalue	% of Variance	Cum %
1	1	9.8224	44.6	44.6
2	2	1.5822	7.2	51.8
3	3	1.0849	4.9	56.8
4	4	.9159	4.2	60.9
5	5	.8562	3.9	64.8
6	6	.7915	3.6	68.4
7	7	.7572	3.4	71.9
8	8	.6685	3.0	74.9
9	9	.6359	2.9	77.8
10	10	.6205	2.8	80.6
11	11	.5287	2.4	83.0
12	12	.5038	2.3	85.3
13	13	.4412	2.0	87.3
14	14	.4108	1.9	89.2
15	15	.4002	1.8	91.0
16	16	.3758	1.7	92.7
17	17	.3376	1.5	94.2

variable	factor	eigenvalue	% of Variance	Cum %
18	18	.3147	1.4	95.7
19	19	.2798	1.3	96.9
20	20	.2624	1.2	98.1
21	21	.2159	1.0	99.1
22	22	.1929	.9	100.0

Table 20

Varimax Rotated Factor Pattern: Depression Check-up,

University Sample (n = 337)

Variables	F.1	F.2	F.3
1 sleep	.12	.73	.30
2 tired	.31	.58	.47
3 slowed	.28	.47	.52
4 appetite	.21	.25	.57
5 constipation	.20	-.22	.73
6 headache	.13	.33	.66
7 muscle tension	.17	.31	.63

Variables	F.1	F.2	F.3
8 sad	.61	.40	.32
9 things not fun	.66	.41	.21
10 self-esteem	.85	.12	.20
11 guilt	.81	.04	.20
12 concentration	.45	.40	.35
13 crying	.69	.15	.32
14 irritable	.41	.37	.38
15 uninterested in people	.58	.48	.13
16 unattractive	.65	.39	.21
17 uninterested in sex	.26	.58	.01
18 low self-esteem	.75	.30	.15
19 world is harsh	.52	.31	.14
20 suicide	.61	.12	.20
21 negative thinking	.68	.35	.09
22 health worry	.38	.57	.23

Interpretation of Findings: Depression Check-up

Based upon most factor analytic practice (Kim & Mueller, 1978), the initial, unrotated statistics for factor analyses were not interpreted. To determine the factor solution, this investigator used both of Thurstone's criteria (cited in Crocker &

Algina, 1986) for simple structure, which is a parsimonious interpretation of the factor structure, and the "criterion of sensibility." "Simple structure" refers to a factor matrix pattern wherein variables should load high on one of the factors and low on other factors. (Loadings less than .30 are considered low and unimportant). Sensibility refers to keeping the number of factors that the researcher wants to interpret to a reasonable level given what the researcher knows about the construct being assessed. Using Thurstone's criterion, the investigator based the factor solution on the Varimax rotation which resulted in 3 factors. The varimax was the easiest to interpret because it satisfied both "simple structure" and "criterion of sensibility."

Factor 1 consisted of all the "emotional symptoms" variables except for "uninterested in sex" which loaded on Factor 2. The variables that had the highest loading on Factor 1 were: sad, things aren't fun, down on self, guilty, concentration, crying, irritable, uninterested in people, unattractive, inadequate, world harsh, suicide, and negative thinking. Uninterested in sex, which was expected to load on Factor 1, loaded on Factor 2. Based on the findings, the investigator decided to label Factor 1 as "Cognitive-Emotional Disturbance."

Factor 2 yielded less clear findings, since there was a hybrid of variables that loaded greater than .50 on this factor. These items were: sleep disturbance, tired feeling, uninterested in sex, and worry about health. (Tired feeling also loaded .47 on

Factor 3). The findings suggested that Factor 2 could be labeled "Psychophysiological Symptoms."

The items that loaded the highest on Factor 3 were items that all tapped physical symptoms: slowed activity/hyperactivity, change in appetite, constipation, headaches, and muscle tension. Contrary to what was expected, sleep disturbance and tired feeling did not load on Factor 3 but loaded .72 on Factor 2. The findings suggested that Factor 3 could be labeled "Physiological Symptoms."

Hypothesis 5-A for the Depression Check-Up was supported for a three-factor structure labeled "Cognitive-Emotional Disturbance," "Psychophysiological Symptoms," and "Physiological Symptoms." Fndings did not support separate factors labeled "cognitive" and "emotional" distress, nor did they support a separation of physical and psychological symptoms. These findings have theoretical implications which will be discussed in Chapter V.

Factor Analysis for the Correa-Barrick Depression Scale

For the Correa-Barrick Depresion Scale, initial, unrotated statistics based on principal components analysis resulted in six factors with eigenvalues greater than 1.0, accounting for 58% of the cumulative variance (Table 21).

Rotated factor analyses using a Varimax (uncorrelated) rotation resulted in seven iterations with six factors with eigenvalues greater than 1.0. Factor 1 had 24 items that loaded .30 or greater. Factor 2 had three items loading greater than .30.

Table 21

Correa-Barrick Depression Scale: Initial Statistics and Eigenvalue
Summary, University Sample (n = 334)

variable	factor	eigenvalue	% of variance	cum %
1	1	10.0121	33.4	33.4
2	2	2.1648	7.2	40.6
3	3	1.6139	5.4	46.0
4	4	1.3353	4.5	50.4
5	5	1.1216	3.7	54.2
6	6	1.0309	3.4	57.6
7	7	.9902	3.3	60.9
8	8	.9121	3.0	63.9
9	9	.8060	2.7	66.6
10	10	.7825	2.6	69.2
11	11	.7664	2.6	71.8
12	12	.7152	2.4	74.2
13	13	.7103	2.4	76.5
14	14	.6696	2.2	78.8
15	15	.6317	2.1	80.9
16	16	.6045	2.0	82.9
17	17	.5275	1.8	84.7

variable	factor	eigenvalue	% of variance	cum %
18	18	.5209	1.7	86.4
19	19	.4947	1.6	88.0
20	20	.4660	1.6	89.6
21	21	.4422	1.5	91.1
22	22	.4142	1.4	92.4
23	23	.3691	1.2	93.7
24	24	.3379	1.1	94.8
25	25	.3314	1.1	95.9
26	26	.3119	1.0	96.9
27	27	.2759	.9	97.9
28	28	.2524	.8	98.7
29	29	.1978	.7	99.4
30	30	.1897	.6	100.0

Factor 3 had three items which loaded greater than .30. Factor 4 had three items load greater than .30. Factor 5 had only two items which loaded greater than .30. Factor 6 had two items (.31 and .47) that loaded greater than 3.0.

Since six factors did not satisfy simple structure, varimax rotation was repeated for five factors. Factor 5 was defined by only one variable; therefore, it was not considered to be

interpretable according to the "criterion of sensibility." Varimax rotation was repeated for four factors. The four-factor structure yielded a simple factor matrix which was easier to interpret and was consistent with theoretical expectation. Based on .30 as the cut-off for important factor loadings, Factor 1 resulted in 13 loadings, Factor 2 resulted in 6 loadings, Factor 3, 6 loadings, and Factor 4, 5 loadings (Table 22).

Interpretation of Findings: Correa-Barrick Depression Scale

Factor 1 had the highest loadings on the following 13 variables: restless, color gray (which also loaded equally as well on Factor 4) worry, depressed, concentration, self-esteem, afraid, irritable, crying, guilt, morbid thoughts, interest in being sensual, and decision-making. Findings suggested that Factor 1 could be labeled "Cognitive-Emotional Disturbance."

Factor 2 had the highest loadings on the following six variables: excited about fun things, notice attractive men/women, look forward to fun things, future outlook, appreciation of sounds, and get excited. Findings suggested that Factor 2 could be labeled "General Outlook".

Table 22

Varimax Rotated Factor Pattern Matrix: Correa-Barrick Depression

Scale, University Sample (n = 334)

Variable	F.1	F.2	F.3	F.4
1 sleep	.14	−.05	.66	.18
2 tired	.11	−.14	.80	.21
3 slowed down	.34	−.20	.75	.21
4 excited	−.09	.81	−.12	−.01
5 trouble with activity	.30	−.39	.43	.28
6 restless	.50	.01	.19	.22
7 wrong with health	.30	−.25	.46	.09
8 lacking color	.45	−.21	.01	.41
9 notice men/women	−.05	.67	.10	−.24
10 look forward to fun	−.16	.86	−.07	−.09
11 brood and worry	.62	−.11	.43	−.05
12 taste not as good	.28	−.24	.21	−.47
13 feel depressed	.56	−.28	.53	.06
14 concentration	.49	−.06	.49	.10
15 future outlook	−.38	.46	−.38	.00
16 self-esteem	.71	−.19	.13	.13
17 afraid	.79	−.06	.02	.17
18 irritable	.60	−.19	.37	.19

Variable	F.1	F.2	F.3	F.4
19 crying	.68	-.15	.25	.26
20 guilt	.77	-.08	.13	.13
21 losing weight	.04	.01	.08	.70
22 dreams	.28	-.07	.15	.44
23 cravings	.06	-.00	.24	.40
24 getting started	.27	-.06	.61	.25
25 sounds	-.06	.45	-.23	-.00
26 excited/happy	-.11	.36	-.16	.26
27 morbid thoughts	.56	-.09	.38	.07
28 loving	.42	-.31	.39	.04
29 decisions	.61	-.10	.18	.04
30 appetite has changed	.23	-.02	.39	.54

Factor 3 had the highest loadings on the following six variables: sleep problems, tired, body feels heavy, trouble with activity, something wrong with health, and trouble getting out of bed. Findings suggested that Factor 3 could be labeled "Physiological Symptoms."

Factor 4 had the highest loadings on the following five variables: taste, losing weight, strange and vivid dreams, appetite change, cravings for sweets and CHO. (It was noted that the

variable on color perception also loaded on Factor 4 at the .41 level.) Findings suggested that Factor 4 could be labeled "Sensory/Perceptual Disturbance."

Hypothesis 5-B for the Correa-Barrick Depression Scale was supported for four factors, but the four factors differed slightly from theoretical expectation. As expected, results supported the factor "physiological symptoms" and another factor "sensory/perceptual disturbance." Consistent with the findings for the Depression Check-up, factor analysis did not support a separate factor structure for cognitive and emotional symptoms for the Correa-Barrick Depression Scale. The theoretical implication of this finding is noteworthy and will be further explored and discussed in Chapter V.

In conclusion, both scales measured a multidimensional construct that seemed to tap broad constructs involving physiological and psychological domains.

Research Question 6

Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate discrimination?

Hypothesis 6-A. The Depression Check-up will discriminate between the university sample and the patient sample. (Partially supported)

Hypothesis 6-B. The Correa-Barrick Depression Scale will discriminate between the university sample and the patient sample. (Partially supported)

Two-sample t-test results were conducted for the Depression Check-up for the university sample and the patient sample. Results showed there were significant differences in the mean depression scale scores for the patient sample (mean = 2.04) versus the university sample (mean = 1.63), t value = 4.95, $p < .01$.

Two sample t-tests were computed for the Correa-Barrick Depression Scale for both the university and patient samples. No significant differences were found.

The t-test analysis was repeated, deleting the six items from the Correa-Barrick Depression Scale that had appeared troublesome to the investigator, based upon the item-total correlations. When the t-test analysis was repeated for the 24-item Correa-Barrick Depression Scale-Short, the results showed significant differences between the patient sample (mean = 3.75) and the university sample (mean = 2.87), t-value = 3.01, $p < .01$.

As a basis for comparison, t-tests were performed on the IDS scale. Results also showed significant differences between the patient sample (mean = .87) and the university sample (mean = .44), t-value = 5.95, $p < .01$.

Additional Analyses

As a final step in analyzing the Depression Check-up and the Correa-Barrick Depression Scale--Short Version, cut-off scores were calculated using normative distributions based on cross tabulation and standard error of measure from the university sample and the patient sample.

Scale Norms

To establish scale norms, several procedures were used. Each procedure for establishing scale norms was repeated exactly for the Depression Check-up and Correa-Barrick Depression Scale--Short Version. The first step involved merging the university sample and the patient sample to form a distribution. Cut-off scores were based upon three criteria: (a) straight percentage from the sample distribution, (b) standard error of measure, and (c) clinical judgement. The first cut-off score was established by calculating a percent of depressed patients based on the basic formula: 50 known depressed patients in a total merged sample of 387 subjects (337 university "Time 1" and 50 patients) = 13% of the total sample. Upon looking at the merged sample distribution, about 87% of the merged sample scored at 2.29 for the Depression Check-up and at 4.52 for the Correa-Barrick Depression Scale. (The 87% figure was used because this represented the balance of the "normal" sample, after 13% of the patients were subtracted from 100%.) Thus, cut-off scores were established at 2.29 and 4.52 for the

Depression Check-up and the Correa-Barrick Depression Scale--Short Version, respectively.

The second cut-off score was established by calculating one standard error of measure (+)1 SEM and (-)1 SEM from the means for the patient sample and the university sample. For the Depression Check-up, the mean score for the patient sample was 2.04, 1 SEM = .09, and the mean score for the normal sample = 1.58 with 1 SEM = .026. The cut-off score for the Depression Check-up was calculated by developing a range of scores, which would represent upper and lower limits of the two sample means using the following method:

depressed mean = 2.04 - 2SEM (.18) = 1.86

normal mean = 2.58 + 2SEM (.052) = 1.63

The values 1.86 and 1.63 were averaged to 1.74, rounded to 1.7, which was selected for the average cut-off score for the Depression Check-up.

For the Correa-Barrick Depression Scale--Short Version, the mean score for the patient sample was 3.75, 1 SEM = .268, and the mean score for the normal sample = 2.87 with 1 SEM = .065. The method for calculating the cut-off score was again repeated using the same method as for the Depression Check-up:

depressed mean = 3.75 - 2SEM (.572) = 3.178

normal mean = 2.87 + 2SEM (.13) = 3.0

3.178 + 3.0 divided by 2 = 3.089, rounded to 3.1 which was the cut-off average score for the Correa-Barrick Depression Scale--Short Version.

The third cut-off score was established by the investigator based upon clinical judgement. For the Depression Check-up, the investigator decided to use the midpoints on each scale (2.5 for the Depression Check-up and 5 for the Correa-Barrick Depression Scale--Short Version), since these scores represent the mid-point between extreme scores on each scale and could reflect the respondent's uncertainty regarding the item.

Finally, the numbers of patients and normals from the university sample who were classified depressed/not depressed were scrutinized to determine how the three different cut-off scores misclassified individuals into either false positive and/or false negatives. The number of individuals who were known depressed patients but were classified as normals was of interest, since this would represent the number of false negatives. In a screening program for depression, a balance of sensitivity (higher false positive and less false negatives) and specificity (less false positive and higher false negatives) are sought. However, sensitivity is usually gained at the expense of specificity, and the screener and/or investigator usually has to make some decisions regarding cut-off scores. Failure to positively screen a clinically depressed individual may have serious consequences in light of the association with suicide and substance abuse. Further, the Depression Check-up and Correa-Barrick Depression Scale were developed for the purposes of screening individuals for

depression, not diagnosing them. The following tables show the
results.

Table 23

Depression Check-up: Depression Results

	depressed (n)	not depressed (n)	total (n)
Classification using 2.29 (percentages) as cut-off score			
patient sample	15	35**	50
university sample	34*	303	337
Classification using 1.7 (SEM) as the cut-off score			
patient sample	33	17**	50
university sample	100*	237	337
Classification using 2.5 (clinical judgment) as the cut-off score			
patient sample	12	38**	50
university sample	21*	316	337

*   false positives
**   false negatives

The average cut-off score of 1.7 for the Depression Check-up seemed the most sensitive; it produced the least number of false negatives but produced the highest number of false positives. It would appear that clinical judgment score of 2.5 might provide the best balance in terms of sensitivity and specificity (Table 23).

Table 24

Correa-Barrick Depression Scale--Short Version: Depression Results

	depressed (n)	not depressed (n)	total (n)
Classification using 4.52 (percentile) as cut-off score			
patient sample	20	30**	50
university sample	35*	302	387
Classification using 3.1 (SEM) as the cut-off score			
patient sample	28	22**	50
university sample	128*	209	337
Classification using 5.0 (clinical judgment) as the cut-off score			
patient sample	17	33**	50
university sample	24*	313	337

*   false positive
*   false negative

Interpretation of Findings: Cut-off Score

Based upon the findings, the Correa-Barrick Depression Scale--Short Version, with a cut-off score of 3.1, produced the least number of false negatives (n = 22) but the highest number of false positive (n = 128). The cut-off score of 5 produced the lowest number of false positives and the second highest number of false negatives (n = 33). It would seem, in the case of the Correa-Barrick Depression Scale, that clinical judgment for an average score of 5.0 seemed the most acceptable.

Discriminant Analysis

Discriminant function analysis was performed on the combined sample (university, "time 1" and patient) using the group classification--university and depressed patient samples--as the dependent variable, and the Depression Check-up, the Correa-Barrick Depression Scale, and the Inventory for Depressive Symptomatology as the independent variables to determine whether the scales discriminated between the two groups.

Results for the Depression Check-up are shown in the Table 25, while results for the Correa-Barrick Depression Scale are shown in Table 26.

Table 25

Discriminant Analysis: Depression Check-up (n = 375)

Source	df	SS	MS	F-ratio	probability level
Model	22	2.687943	.1221792	1.14	.300*
Error	352	37.669390	.1070153		
Total	374	40.35733	.1079073		

*   Not signficant

Table 26

Discriminant Analysis: Correa-Barrick Depression Scale (n = 352)

Source	df	SS	MS	F-ratio	probability level
Model	30	3.799609	.1266536	1.10	.332*
Error	321	36.924820	.1150306		
Total	351	40.724429	.1160243		

*   Not signficant

Another procedure was completed to determine if any of the
items on either the Depression Check-up or the Correa-Barrick

Depression Scale would discriminate the patient sample from the university sample. Summary statistics for the university and depressed patient sample and item-by-item t-values and probability levels in the merged sample (combined patient and university samples, n = 387) were performed for the Depression Check-up (Table 27) and the Correa-Barrick Depression Scale (Table 28).

There were three items on the Depression Check-up that discriminated between the depressed and non-depressed sample: item #5 ("Constipation"), t-value = 2.05, p < .05, and item #15 ("Uninterested in people"), t-value = -2.04, p < .05, and item #18 ("Thinking that something is wrong with you"), t-value = 2.3, p < .05.

Three items on the Correa-Barrick Depression Scale discriminated between the depressed sample and the university sample: item #10 ("I look forward to fun things (hobbies/interests/social contact) as much as I used to"), t-value = -2.8, p < .01; item #15 ("I have much to look forward to in the future"), t-value = 2.4, p < .05, and item #20 ("I feel responsible for bad things that have happened"), t-value = 2.5, p < .05.

Table 27

Depression Check-up: Summary Statistics for the University and
Depressed Samples and *t*-value and Probability Levels for the
Combined Sample (n = 387)

	Summary Statistics							
	University			Depressed			Combined	
Item	n	mean	SD	n	mean	SD	*t*-value	*p*
1	336	1.99	.89	49	2.40	.79	.76	.44
2	335	2.02	.77	49	2.40	.95	−0.71	.47
3	336	1.58	.73	49	2.14	.88	.41	.68
4	336	1.36	.68	50	1.90	.99	−1.15	.25
5	336	1.29	.57	50	1.70	.88	2.05	.04*
6	336	1.92	.84	50	2.34	.92	.93	.35
7	336	1.86	.83	50	2.22	.93	−1.15	.25
8	336	1.64	.79	50	2.16	.87	−0.75	.45
9	336	1.54	.75	50	2.22	.88	.97	.33
10	336	1.42	.75	50	2.13	1.00	−0.22	.82
11	336	1.29	.63	50	1.72	.97	.33	.74
12	335	1.60	.65	50	2.30	.99	.53	.59
13	336	1.29	.60	50	1.78	.84	1.12	.26
14	336	1.64	.66	50	1.98	.74	−0.26	.79
15	333	1.41	.69	49	1.96	.79	−2.04	.04*
16	335	1.60	.77	50	1.92	.99	−0.44	.66
17	335	1.73	.89	49	2.25	1.04	−1.32	.18
18	335	1.51	.76	50	2.20	1.04	1.73	.08
19	336	1.65	.79	50	1.78	.78	−1.01	.31

(table continues)

Summary Statistics

Item	University			Depressed			Combined	
	n	mean	SD	n	mean	SD	t-value	p
20	335	1.08	.36	50	1.36	.62	.29	.77
21	335	1.53	.69	50	1.99	.81	.97	.33
22	336	1.74	.82	50	2.08	.87	.44	.66

Minimum item score = 1

Maximum item score = 4

* $p < .05$.

Table 28

Correa-Barrick Depression Scale: Summary Statistics for the University and Depressed Samples and *t*-value and Probability Levels for the Combined Sample

	Summary Statistics							
	University			Depressed			Combined	
Item	n	mean	SD	n	mean	SD	*t*-value	*p*
1	335	3.35	2.94	48	3.63	2.94	−.8	.43
2	333	2.30	2.32	48	4.15	2.84	.3	.79
3	336	2.77	2.53	49	4.47	2.76	−1.4	.16
4	336	7.33	2.35	48	4.91	2.94	1.2	.24
5	336	1.36	1.67	49	3.62	2.46	.5	.64
6	335	1.68	1.94	49	3.36	2.52	.8	.42
7	334	1.88	2.32	48	4.16	3.29	.1	.95
8	335	.79	1.30	49	2.29	2.23	.7	.49
9	335	7.66	2.45	47	3.93	3.03	1.0	.33
10	335	7.82	2.27	49	4.40	3.04	−2.8	.00*
11	335	3.56	2.77	49	4.90	2.96	−1.1	.26
12	335	1.45	1.93	49	2.82	2.94	.6	.57
13	335	2.09	2.39	49	3.96	2.93	1.0	.34
14	336	2.89	2.46	43	4.38	3.19	.9	.37
15	335	7.65	2.42	50	3.74	2.97	2.4	.02**
16	335	2.20	2.07	50	3.88	2.81	.2	.85
17	335	1.34	1.94	50	2.86	2.49	.5	.62
18	336	2.16	2.28	50	3.17	2.67	.2	.85
19	335	1.48	2.09	50	2.55	2.66	.3	.78

(table continues)

Summary Statistics

Item	University			Depressed			Combined	
	n	mean	SD	n	mean	SD	t-value	p
20	333	1.93	2.37	50	3.09	2.73	2.5	.01**
21	335	.68	1.45	50	1.03	1.80	1.1	.26
22	336	1.30	1.87	50	2.99	3.21	1.5	.13
23	335	2.58	4.32	50	4.09	3.40	.9	.34
24	335	2.85	2.85	50	4.74	3.37	−1.1	.26
25	336	7.86	2.52	50	3.20	2.63	1.6	.11
26	336	8.40	4.24	49	3.58	2.61	.3	.78
27	334	2.23	2.24	50	2.88	2.65	.5	.58
28	327	2.65	2.82	50	4.20	3.18	.4	.71
29	336	2.77	2.60	50	4.55	3.50	.7	.50
30	336	2.24	2.60	50	4.07	3.25	.7	.50

Minimum item score = 0
Maximum item score = 10

* $p < .01$.
** $p < .05$

Interpretation of Findings

Even though the t-tests showed that the mean depression scores between the depressed and non-depressed sample were significantly different for the Depression Check-up and Correa-Barrick Depression Scale--Short Version, discriminant analysis failed to show that the Depression Check-up, the Correa-Barrick Depression Scale, and the Inventory for Depressive Symptomatology would discriminate between the two groups. This finding was expected for the Correa-Barrick Depression Scale but unexpected for the Depression Check-up. Since the two-sample t-tests from Table 23 showed significant differences for the Depression Check-up, it was expected that the Depression Check-up would discriminate.

One possible explanation for this null finding could be because the mean depression scores in the university and patient samples did not differ as greatly in magnitude as expected. Since the patient sample had been under treatment, the patient scores would tend to become more "normalized." When reviewing the mean depressed scores for the Depression Check-up from Table 2, the mean depression score for the Depression Check-up was 1.58 for the university sample and 2.04 for the patient sample. Although the mean Depression Check-up score differences were statistically significant, they did not appear impressive from a practical standpoint.

Summary

Results of the data analysis were presented in Chapter IV. Data were presented in statistical tables in order to provide a summary of the psychometric performance of the Depression Check-up and the Correa-Barrick Depression Scale based upon classical test theory and comparisons to other known published depression scales.

A narrative and statistical summary of the study, findings, conclusions, implications, and recommendations for further study will be discussed in Chapter V.

CHAPTER V

SUMMARY, CONCLUSIONS, IMPLICATIONS, AND
RECOMMENDATIONS FOR FUTURE RESEARCH

The purpose of this chapter is to summarize the research findings, present conclusions, describe limitations of the study, and address the implications for health education practice, theory, and research. Finally, directions for future research will be proposed.

Review of the Research Questions

The objective of this investigation was to conduct a psychometric assessment of two new depression self-rating scales, the Depression Check-up (Schiraldi, 1987) and the Correa-Barrick Depression Scale. There were six research questions:

1. Will the Depression Check-up and the Correa-Barrick Depression Scale possess convergent validity?

2. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate concurrent validity?

3. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability based upon internal consistency?

4. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability over time?

5. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate construct validity?

6. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate discrimination?

Review of the Research Design

A survey instrument consisting of demographic data, the Depression Check-up, the Correa-Barrick Depression Scale, and the Inventory for Depressive Symptomatology, Self-Report (Rush et al., 1986) was designed to answer the above research questions. A survey research design was used to collect data from Sample I which was composed of 1,200 faculty and staff employed at a metropolitan comprehensive university and from a subset sample of Sample I (hereafter called "Sample I Subset"). This subset consisted of the respondents who returned the survey after a second mailing. Since the investigator was interested in determining reliability over time of the Depression Check-up and the Correa-Barrick Depression Scale and, in order to compute test-retest correlation coefficient between Sample I and the Sample I Subset, identical surveys were mailed at two different times (labeled "Time 1" and "Time 2"). Sample II was composed from 200 depressed outpatients. Return rates for Sample I and Sample II were 28% and 25%, respectively.

Summary of the Findings

The following section summarizes the demographic characteristics for Sample I and Sample II and the answers to each of the six research questions. A summary of the psychometric results of the study was also developed (Appendix J).

Findings from Demographic Variables

From Sample I, 337 faculty and staff responded to the survey instrument. Demographic characteristics for Sample I were: mean age 46 years, 57% females, 53% married, 87% Caucasian, 33% doctorally prepared, and an equal composition from among faculty, professional staff, and staff. About 8% of the university sample reported being currently depressed which was almost twice the 4.4% prevalence rate of depression reported by Weissman et al. (1988).

From the Sample I Subset 203 faculty and staff responded to the second survey instrument. Demographic characteristics for Sample I Subset were: mean age 46, 57% female, 55% married, 89% Caucasian, 31% doctorally prepared, and an equal composition from among faculty, professional staff, and staff. About 5% of the sample reported being currently depressed.

From Sample II, 50 depressed persons under treatment in an outpatient setting from Sample II responded to the survey instrument. Demographic characteristics for Sample II were: mean age 46 years, 63% females, 54% married, 96% Caucasian, and 26% with some college, 52% with a baccalaureate or graduate degree, and 6%

doctorally prepared; 33% reported current depression. Demographics for the patient population supported the higher prevalence of depression reported by Feighner and Boyer (1991) in the female population. In summary, Sample I, Sample I Subset, and Sample II represented a highly educated, middle-aged, Caucasian population. The main difference among the three samples was reflected in the reported prevalence of current depression.

Findings for the Research Questions

Research Question 1. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate convergent validity with two known published scales: the Beck Depression Inventory and the Inventory for Depressive Symptomatology, Self-report?

Research findings from both the pilot and main studies supported this research question. The findings supported the validity of both the Depression Check-up and the Correa-Barrick Depression Scale for measuring depression in a university and depressed patient sample based on the validity coefficients that the Depression Check-up and Correa-Barrick Depression Scale demonstrated with the Beck Depression Inventory (BDI) and the Inventory for Depressive Symptomatology, Self-report (IDS-SR).

The findings showed higher validity coefficients for the Depression Check-up and the Correa-Barrick Depression Scale with the IDS-SR than with the BDI. According to Lambert, Christensen and DeJulio (1983), one explanation for this finding may be that

the Beck Depression Inventory did not include all of the physical symptoms from the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DMS-III-R; 1987). The Inventory for Depressive Symptomatology, the Depression Check-up, and the Correa-Barrick Depression Scale were all composed to include physical symptomatology as specified by the DSM-III-R.

Research Question 2. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate concurrent validity with the variables that are expected to correlate with depression?

The hypothesis for Research Question 2 was only partially supported for the Depression Check-up and the Correa-Barrick Depression Scale. Significant group differences were not found for either the Depression Check-up or the Correa-Barrick Depression Scale for the variables "history of depression" and "taking medication for depression" in the university sample. However, significant group differences were consistently found on the Depression Check-up and the Correa-Barrick Depression Scale for the variables "family history" and "presence of a current depression." Significant validity coefficients were also found for the scale item on depression which asked about level of depression and the total depression scores from the scales.

Research Question 3. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability based upon internal consistency?

The Depression Check-up and the Correa-Barrick Depression Scale showed high reliability coefficients for internal consistency based upon the criterion that a reliability correlation coefficient $r = .90$ was high.

Research Question 4. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate reliability over time?

For the Correa-Barrick Depression Scale, the stability coefficient $r = .70$ was not as high as the Depression Check-up $r = .81$. Even after eliminating six of the troublesome scale items, the Correa-Barrick Depression Scale--Short Version still failed to rise above a reliability coefficient of $r = .73$. Results for the Correa-Barrick Depression Scale were comparable to the stability coefficient of the Inventory for Depressive Symptomatology, $r = .71$, in the university sample. Although the result for the Correa-Barrick Depression Scale was comparable to the Inventory for Depressive Symptomatology, the finding was lower than expected.

One explanation for the lower result may be due to problems with respondent memory recall which was reported by Gift (1989). One other explanation could be respondent fatigue, but this explanation was discounted because the Correa-Barrick Depression Scale was formatted as the first scale on the survey.

Test-retest reliability coefficients for the patient sample were unknown, as the investigator did not conduct this procedure for two reasons: (a) patient cooperation in completing surveys was a serious barrier to the research study, and (b) test-retest over a

two-week period was questionable for a depressed sample. Hedlund and Vieweg (1979) reported difficulties with test-retest in a patient population due to the fluctuating nature of depression, especially in a patient sample under treatment, and recommended that test-retest be done at two- to three-day intervals.

Research Question 5. Will the Depression Check-up and the Correa-Barrick Depression Scale demonstrate construct validity?

Construct validity, based upon factor analysis, supported a three-factor structure for the Depression Check-up and a four-factor structure for the Correa-Barrick Depression Scale. The study did not support the hypothesized labels for each of the factor structures, but the factor structure findings from this study were conceptually similar to other construct findings from the literature. Factor analysis for the Beck Depression Inventory revealed three factors: "negative attitudes," "performance difficulties," and "somatic complaints" (Beck, Steer & Garbin, 1988). Development of the Zung Self-rating Depression Scale (Zung, 1965) was based on conceptualization of three factors: "pervasive affect," "physiological equivalents," and "psychological concomitant." The Inventory for Depressive Symptomatology revealed four factors: "mood and cognition," "vegetative factor," "atypical symptom factor," and an "anxious depression factor" (Rush et al., 1986).

All of these published scales showed a one-factor structure that could be labeled "physical symptoms." Factor structure for

the Depression Check-up and the Correa-Barrick Depression Scale
supported a similar one-factor structure for physical symptoms.
Factor analysis on the Correa-Barrick Depression Scale also
suggested a factor labeled "sensory perceptual disturbance." Even
though alterations in color perception, such as depressed persons
perceiving shades of gray and black, were reported by Goodwin and
Jamison (1990), this variable was not included as an item on any of
the reviewed depression scales nor was it reported in any of the
factor analytic studies reviewed by the investigator. This
suggested that a fourth factor, "sensory/perceptual disturbance,"
was a new factor finding.

Research Question 6. Will the Depression Check-up and the
Correa-Barrick Depression Scale demonstrate divergent validity?
Findings supported the research question for the Depression Check-
up but not for the Correa-Barrick Depression Scale. Since there
had been six troublesome items on the Correa-Barrick Depression
Scale, it was thought that if these items were deleted, the Correa-
Barrick Depression Scale--Short Version would discriminate between
the patient and the university sample and thus support divergent
validity. The investigator's expectation was met. When analysis
for the Correa-Barrick Depression Scale--Short Version was
repeated, significant differences in mean scores between the
patient and university samples emerged.

One unexpected finding was that neither the Depression Check-up nor the Correa-Barrick Depression Scale discriminated between the university and patient groups.

Conclusion

Validation of the 30-item Correa-Barrick Depression Scale was more troublesome than the Depression Check-up. Although the Correa-Barrick Depression Scale demonstrated validity and reliability in the patient and university samples, results for the Correa-Barrick Depression Scale--Short Version were slightly better than the original Correa-Barrick Depression Scale, notably on convergent validity and test-retest reliability.

Based upon the findings from this investigation, there was evidence to support an inference that the Depression Check-up and the Correa-Barrick Depression Scale demonstrated initial validity and reliability for measuring depression in a university sample and a known, depressed patient sample. This conclusion was based upon two criteria: (a) performance of the Depression Check-up and the Correa-Barrick Depression Scale at acceptable psychometric standards based upon the classical test theory (criterion approach) of Crocker and Algina (1986) and (b) comparisons with two published scales, the Beck Depression Inventory and the Inventory Depressive Symptomatology, Self-Report (normative approach).

The Depression Check-up and the Correa-Barrick Depression Scale failed to demonstrate discrimination between the university

and depressed patient samples. One factor that might have accounted for this null finding was the insufficient sample size of newly diagnosed and/or untreated patients who would be expected to show higher depression scores. One other problem could have been that each sample group was fairly homogenous with respect to depression.

Respondent preference toward the visual analog scale or the four-point scale emerged during the qualitative phase of the study. Findings concerning respondent preference, based upon the investigator's interviews of the university and patient samples, were inconsistent. Previous research findings reported that respondents preferred the simplistic format of a visual analog scale (Aitken, 1969). Findings from this study suggested that level of education and emotional state influenced respondent preference for a scale format. Specifically, depressed patients preferred the visual analog scale since it was the least taxing cognitively, and the university sample preferred the four-point categories because of its emphasis on vocabulary.

Scale validation is a continuous process, and the results from this study were viewed as preliminary findings. Prior to implementation of the scales in a clinical setting, further replication of the study in a different sample would be preferred. It should be noted that the university sample used in this study represented a highly educated adult sample and was not representative of the general population. Therefore, results of

this study cannot be generalized beyond the characteristics of the two samples used in this investigation.

One interesting finding in this study was the relationship between the sensory-perceptual items (taste and color perception) and clinical depression scores; in particular, the item on color perception showed a high positive correlation ($r = .77$) with clinical depression scores in the patient sample. These findings may have reflected a subtype of depression, an additional variable related to clinical depression, or happenstance. It is important to consider further replication and investigation into this area since alterations in color perception may be a biological marker for clinical depression. At the present time, any conclusion about the relationship between sensory-perceptual disturbance and clinical depression has been made tentatively.

It would be interesting to determine how results may have been affected if the patient sample had been more depressed; that is, if their depression scores reflected a greater severity of depression. According to the descriptions, 33% of the patient sample reported they were depressed; 96% of this group were taking medication to treat depression. In contrast, 8% of the university sample reported they were depressed, and 95% reported they were not taking medication to treat depression. If the patient sample had a higher percent of patients reporting current depression, one could conjecture that the Depression Check-up and the Correa-Barrick Depression Scale might demonstrate better discrimination results

between the depressed sample and the university sample. The mean cut-off score for depression might shift to a higher value since an untreated patient sample would be expected to show much higher depression scores.

Limitations of the Study

There were several limitations to this study. Firstly, the study's Sample I and Sample II represented a convenience sample, and the demographic data indicated that the depressed sample and the university sample were well educated. Results of this study should not be generalized beyond these two populations.

Secondly, social desirability may have influenced subject responses, especially since depression is a sensitive topic. Problems with social desirability on self-report scales had been reported by DeVellis (1991), and it was considered by the investigator to be a factor in this study. Specifically, the range of depression scores were probably restricted because respondents from the university sample did not want to disclose honest feelings, and the patient sample may not have.

Thirdly, a return rate of 28% for the university sample and 25% for the patient sample was slightly below the investigator's expected response rate of 30% Nevertheless, the return rate of 28% for Sample I equaled 337 returned surveys which met the requirement for the ratio of numbers of subjects (five to ten) per scale item for factor analysis (DeVellis, 1991) for the 30-item

Correa-Barrick Depression Scale and the 22-item Depression Check-up. Since the return rate for mailed questionnaires is usually in the 10% to 50% range (Kidder, 1981), this study was within the norm for response rate. Since follow-up mailing to non-respondents was not possible due to the anonymity of the study, information about the subjects who did not respond remained unknown. This constrained the representativeness of the population samples studied, since, according to Kidder, non-respondents may be different; that is, often less educated and less interested in the survey topic than respondents.

Fourthly, since the depressed sample was under treatment, total depression scores for the patient and the university sample did not practically differ. An untreated depressed sample would be expected to yield a wider range in scores. If so, this might have changed results for the discriminant analysis from null to positive group discrimination. Since it is not ethical to withhold treatment from depressed patients, this method could not and should not be implemented.

Implications for Health Education

Findings in this study have implications for health education practice, theory, and research.

Health Education Practice

The Depression Check-up and the Correa-Barrick Depression Scale may be utilized in screening programs for adult populations in which the goal is to screen, not diagnose, individuals for possible depression. According to Mausner and Kramer (1985), "a screening test should provide a good preliminary indication of which individuals actually have the disease and which do not" (p. 217). There are two types of validity in screening tests: sensitivity and specificity. Sensitivity was defined as "the ability of a test to correctly identify those who have the disease" (p. 217). Specificity was defined as "the ability of a test to identify correctly those who do not have the disease" (p. 217). Mausner and Kramer summed up the problems in screening:

> An ideal screening test would be 100 percent sensitive and 100 percent specific. In practice this does not occur; sensitivity and specificity are usually inversely related. That is, one usually achieves high sensitivity at the expense of low specificity, and vice versa. (p. 217)

Decisions about cut-off scores should be made carefully with consideration given to the fact that the lower the cut-off score, the probability of finding more depressed cases (sensitivity) is maximized. According to the principle of specificity as mentioned by Mausner and Kramer (1985), a higher cut-off depression score should be utilized if the screener's goal is to minimize the prevalence of false positives.

Care must be taken when interpreting cut-off scores for the Depression Check-up and the Correa-Barrick Depression Scale, since

this study was performed on a highly educated adult population.
Beck, Steer, and Garbin (1988) reported that the appropriateness of
cut-off scores is variable depending on the sample and the purpose.
The health educator who uses the Depression Check-up and the
Correa-Barrick Depression Scale should understand that the scales
are instruments for measuring severity of depression as it relates
to a state condition, not a trait, in an adult population.
Although the study suggested that the Depression Check-up and the
Correa-Barrick Depression Scale were reliable over time in a non-
depressed sample, scores may vary widely within a 24-hour period.
In addition, the scales could also be measuring depression caused
by a stressful condition; for example, death of a significant
person. In this case, scores on the depression scale might
temporarily produce inflated scores.

Since health education screening programs are designed for
the purpose of detecting the possibility of a condition and do not
constitute diagnostic evaluations, it is recommended that
sensitivity, using higher cut-off scores, be a guide when screening
for depression. However certain caveats apply. Health educators
must be responsible and prudent in implementing a depression
screening program. Nearby resources and facilities or the presence
of a psychologist/psychiatrist on site must be available for
individuals who are seriously depressed/suicidal/dysfunctional. A
severely depressed individual should be escorted to a facility and
a follow-up telephone call to the facility by the screener should

be made to insure that the participant is under care. Moreover, careful counselling must be undertaken to assure individuals who score above the established cut-off that results may be due to current stressors or to screening error and not due to a clinical depression. Brief, on-site stress management counselling, literature, and resources should be offered for individuals experiencing current stressors. It should be emphasized to participants that the depression tests are only a screening device and not a diagnostic tool.

Mental health screening is not as clear-cut as blood pressure screening, since the instrumentation for blood pressure is standardized and follow-up measurements are easier to conduct. Whereas, in a mental health screening, individuals who score above the cut-off should be referred to a mental health facility, since they may be at risk for a clinical depression, may have a clinical depression, or may be symptomatic due to stressors. Names, addresses, and telephone numbers of facilities and/or resources should be offered to participants for referral. A follow-up evaluation that includes a psychiatric interview, diagnosis, and treatment may be done, and participants should be so advised.

The Depression Check-up and the Correa-Barrick Depression Scale may also be used by the health educator for educational purposes. Items listed on the scales are useful handouts for instructing persons about the signs and symptoms of depression and can be used as self-assessment guides.

Health Education Theory

Of considerable interest to health education is the finding that neither the Depression Check-up nor the Correa-Barrick Depression Scale showed a separate factor structure for emotional and cognitive symptoms. Specifically, emotional and cognitive symptoms both loaded on one factor. This was a very interesting finding, since it lent support to cognitive theory (Ellis, 1973; Beck, 1976; Everly, 1990). Cognitive theorists believed that emotions and thought processes were inseparable and that, if an individual had negative thoughts, then that individual would start to feel depressed. Disputing irrational beliefs was the basis for rational-emotive therapy (Ellis, 1973). However, the logic of cognitive therapy may also become circular; conversely, an individual may start feeling depressed and then start thinking negative thoughts. Therefore, it is methodologically complex to determine whether distorted negative thoughts trigger the depression or whether the depression triggers the negative thoughts. Regardless of the perspective, findings from this study lend further support to viewing depression in a paradigm that examines both cognitive and affective states simultaneously, not as dichotomous entities.

Another finding from this study was the importance of psychological variables in depression. Some of the biological items (appetite, weight loss) showed weaker correlations with depression than did the psychological variables. This tended to

support the view of Steer and Beck (1985): that biological items may not always be present in major depression. This has implications for future development of the diagnostic criteria for depression. It may be that biological items are not sufficient indicators for diagnosing depression. In fact, biological items may also mimic emotional and or physical problems other than depression.

Health Education Measurement and Research

Future research should be undertaken in health education to develop multiple-item visual analogue scales to measure health education theories, constructs and models such as health belief, compliance, and stress. In today's economically oriented health care system, emphasis needs to be placed on evaluation research to determine and document the efficacy of health education programs. Finally, the health education discipline needs to incorporate more emphasis on measurement and psychometrics as a sub-specialty to advance the field.

The Visual Analog Scale may prove to be sensitive to changes for evaluating a health education program or protocol. Current four-point Likert-type scales may not be sufficiently fine-tuned to detect subtle, yet practical, differences. Clearly, a more sensitive tool is desirable especially for measuring outcomes because program funding needs to be justified in these times of budget constraints. The trend today in evaluation is toward

program evaluations that are "outcome based." Specifically, this involves measuring changes in patient attitudes, beliefs, and feelings following a health education intervention. The visual analog scale can also be used to test consumer satisfaction with a program, as consumer satisfaction is fast becoming an important litmus test for evaluating program effectiveness. This may be especially true in situations where outcome variables may be difficult to quantify and measure. Green and Lewis (1986) defined evaluation as a "comparison of an object of interest against a standard of acceptability" (p. 171). In health education, the primary purpose for evaluation is accountability for services rendered, and health education programs can be evaluated for accountability at three levels: formative, impact, and outcome. In formative evaluation, the evaluator assesses appropriate content, methods and performance. Impact evaluation focuses on the immediate impact the program has on knowledge, beliefs, attitudes, skills, social supports, and behavior. Outcome evaluation measures factors such as incidence and prevalence of risk factors, morbidity, and mortality (Green & Lewis, 1986). Outcome evaluation usually requires a larger sample and a longitudinal study.

In evaluating patient education and stress management health education programs for depression, each type of evaluation can be done. In a formative evaluation, for example, a multiple-item visual analog scale can be used by peer review health education specialists to rate the quality of the program. Similarly, program

participants can be asked to complete a single-item visual analog scale with the following anchors: "Most satisfied I have been with a program" to "Least satisfied with the program."

The health educator may use the visual analog scale for an impact evaluation of a program on stress management. A depression visual analog scale can be used to assess symptom severity, such as "Most depressed I have ever been" to "least depressed I have ever been." Beliefs about treatment using a visual analog scale may be valuable in predicting a person's intention to comply, for example, "controlling stress is important in preventing depression" and "taking medication and seeing my therapist is important in treating depression."

Ultimately, outcome evaluation can assess morbidity, and morbidity can be assessed by public health statistics. Disability is a dependent variable that seems an acceptable and practical measure of the effectiveness of health education programs. It can be measured by a single- or multiple-item visual analog scale; for example, "the best I have ever functioned" versus "the worst I have ever functioned."

Suggestions for Future Research

There are a number of research topics suggested by the results of this study. Based on this study's findings, suggestions for future research are:

1. That clinical research on the relationship between sensory items (color perception and taste) and clinical depression be undertaken in a depressed patient population.

2. That studies on test-retest reliability over a two- to three-day interval for the Depression Check-up and the Correa-Barrick Depression Scale be undertaken in a clinically depressed patient population. It was recommended by Hedlund and Vieweg (1979) that test-retest be conducted at two- to three-day intervals due to the fluctuating nature of depression.

3. That the sensitivity of both the Depression Check-up and the Correa-Barrick Depression Scale be evaluated by using simultaneous depression self-rating scales and clinical rating to determine their merit as a sensitive evaluation outcome measure for either stress management or mental health settings.

4. That the sensitivity of the Visual Analogue Scale, using the multiple-item Correa-Barrick Depression Scale, be evaluated in a clinically depressed population to assess whether it is superior to the four-point depression scales for detecting clinical changes in depressed persons.

5. That the validity and reliability of the Depression Check-up and the Correa-Barrick Depression Scale be replicated in other non-depressed and depressed samples. Consideration should be given (a) to administering the scales one at a time per patient visit in order to maximize response rate and (b) to administering the scale to patients who are either in an inpatient facility or

have been recently diagnosed to achieve a greater range of scores. Hopefully, better discrimination might be illustrated by the Depression Check-up and the Correa-Barrick Depression Scale.

6. That further study be undertaken concerning how educational level and emotional state may influence a respondent's preference for scale format.

7. That research be undertaken relevant to developing and validating multiple-item visual analogue scales for measuring health education constructs such as stress, compliance, health beliefs and behavior.

APPENDICES

Appendix A

Depression Check-up Scale

RATE YOURSELF

	NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
SCALE I. BODILY SYMPTOMS				

During the last week have you experienced...

	NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
A. Sleep disturbance (trouble getting to sleep or staying asleep, early wakening, or sleeping longer than usual)	1	2	3	4
B. Tired feeling (no energy)	1	2	3	4
C. Slowed activity (sluggish movement or speech; hard to get moving) OR Hyperactivity (excessive movement/working/fidgetting)	1	2	3	4
D. A change in appetite, weight, or the amount you eat (a decrease or increase that you did not plan)	1	2	3	4
E. Constipation	1	2	3	4
F. Headaches, or other aches and pains	1	2	3	4
G. Muscle tension	1	2	3	4

SCALE 1 (BODILY SYMPTOMS) TOTAL _____
(Add numbers circled)

SCALE II. EMOTIONAL SYMPTOMS

During the last week how often have you felt...

	NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
A. Sad (gloomy, discouraged, blue, numb, empty, or like you just don't care)	1	2	3	4
B. That things aren't much fun anymore	1	2	3	4
C. Down on yourself (worthless, low self-esteem, unlovable, self-disliking)	1	2	3	4
D. Guilty (like a bad person)	1	2	3	4
E. Unable to concentrate, think clearly, remember, or make decisions	1	2	3	4
F. Like crying (or that you would cry, but no longer seem able to)	1	2	3	4
G. Irritable (touchy, nervous, jittery)	1	2	3	4
H. Uninterested in people	1	2	3	4
I. Unattractive	1	2	3	4
J. Uninterested in sex	1	2	3	4

SCALE III (THOUGHT SYMPTOMS)   TOTAL _____
(Add numbers circled)

SCALE III. THOUGHT SYMPTOMS

During the last week have you found yourself...

	NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
A. Thinking that something is wrong with you (you're inadequate-lacking something needed to be happy or successful)	1	2	3	4
B. Thinking that the world and its people are harsh and unfair	1	2	3	4
C. Thinking of suicide (either planning it, thinking you'd like to, or thinking it would be better if you weren't around)	1	2	3	4
D. Noticing mainly the negative aspects of situations or people	1	2	3	4
E. Worrying about your physical health or body functions	1	2	3	4

SCALE III (THOUGHT SYMPTOMS)   TOTAL _____
(Add numbers circled)

GRAND TOTAL (Sum of totals from Scales I, II, and III) _____

Appendix B

Correa-Barrick Depression Scale

[First Cover Letter to Faculty]

Dear Faculty and Staff Member:

I am conducting research under the auspices of a faculty research grant and am also completing a doctorate at the University of Maryland. To graduate, I need your help in completing a "Confidential Survey." The FIRST survey is enclosed. The SECOND survey will be administered in about two weeks. The purpose of this study is to determine the validity and reliability of the enclosed survey.

To thank you for your participation in this study, you are eligible to enter your name in a random drawing for a $100.00 cash prize after you have mailed the completed SECOND survey.

The survey will take about 15 minutes to complete. After you are finished place in the Interoffice envelope, send through campus mail to "C. Barrick, XXXXX Department. Please do NOT write your name on the survey. The survey is CONFIDENTIAL and ANONYMOUS.

Participation in this study is voluntary. Your employer will NOT be given any information about individual responses. Your employment will NOT be affected if you chose not to participate. You do not have to answer any question(s) you prefer not to.

If you have any concerns about any of your responses to any of the questions, please contact one of the following: XXXXX Counseling Center at [telephone number]; your primary physician or health care provider; the local County Mental Health Department (listed in the government section of the telephone book, or call 411); and XXXXX Medical Institutions for referral information at [telephone number].

I thank you for your cooperation. If you have any questions about the study, please feel free to call me at [telephone number] or XXXXX, Chairperson, Institutional Review Board at [telephone number].

Thank you,

Christina B. Barrick, MS, RN
Assistant Professor
[academic department]
[college/university]

[Second Cover Letter to Faculty and Staff]

Dear Faculty/Staff:

Thank you for responding to the First "Confidential Survey." This SECOND "CONFIDENTIAL SURVEY" is still part of my research, and I need your help again.

To participate in a random drawing to win a cash prize of $100.00, please complete and mail the enclosed survey to "C. Barrick, XXXXX" using the Interoffice mailer ON or BEFORE SEPTEMBER 10, 1993. The drawing will be held September 1993 and will be conducted by the research investigator AND NOT BY ANY PARTICIPANT. This survey should take about 10 to 15 minutes to complete. Your timely response is important to the quality of this study. Your cooperation is greatly valued.

The survey is confidential and anonymous, and your participation is voluntary. If you have any questions, please feel free to call me or XXXXX [telephone number].

Thank you,

Christina B. Barrick, MS, RN
[academic department]
[telephone number]

. . . .

For participants to remain eligible to win the cash prize, the SECOND survey must be completed and returned by campus mail on or BEFORE SEPTEMBER 10, 1993. Please cut at the dotted line, and send in Interoffice campus mail to "C. Barrick, XXXXX."

What do you think this study was about?

_____

Your Name : _____

Address: _____

_____

_____

Code #:_____ (For verifying returns of both surveys only.)

If you would like a summary of the findings, please enclosed a self-addressed envelope to "C. Barrick, XXXXX."

"CONFIDENTIAL SURVEY"

Study Conducted by:

Christina B. Barrick, MS, RN
Assistant Professor, [academic department]
[college/university]
Ph.D. Candidate, University of Maryland
College Park, MD

September 22, 1993

Dear Faculty/Staff:

Thank you for responding to the first "Confidential Survey". The enclosed, second "Confidential Survey" is still part of my research, and I need your help in completing it again even though the questions are the same as the first "Confidential Survey".

To be eligible to participate in a random drawing to win a cash prize of $100.00, <u>PLEASE COMPLETE AND MAIL THE ENCLOSED SECOND "CONFIDENTIAL SURVEY" BEFORE OR ON WEDNESDAY, SEPTEMBER 29, 1993 TO "C. BARRICK, [ACADEMIC DEPARTMENT]</u> in the campus mail. The survey should take abut 15 minutes to complete. Your timely response is very important to the quality of this study.

The recipient of the cash prize will be contacted by telephone no later than Friday October 29, 1993. The drawing will be randomly conducted by the investigator.

The survey is confidential and anonymous. Your participation is voluntary. Individual responses will not be given to your employer. If you have any questions, please feel free to call me at ext. xxxxx.

Thank-you,

Christina Barrick

**** IMPORTANT ****
_____

TO ENTER THE DRAWING, PLEASE COMPLETE AND RETURN THE INFORMATION BELOW, ALONG WITH YOUR SURVEY, IN CAMPUS MAIL TO "C. Barrick, [academic department], by <u>WEDNESDAY SEPTEMBER 29, 1993</u>. THE DEADLINE FOR RECEIVING SURVEYS IS OCT. 1, 1993. SURVEYS RECEIVED AFTER OCT. 1 ARE NOT ELIGIBLE FOR THE DRAWING BUT MAY STILL BE USED IN THE STUDY. (To maintain anonymity, I will separate the ENTRY FORM below from your survey responses).

What do you think this study was about? _____

Your name: _____

Telephone #: _____

Campus Address: _____

If you would like a summary of the findings, please enclose a self-addressed envelope to "C. Barrick, [academic department]."

CONFIDENTIAL SURVEY

PART I   BACKGROUND INFORMATION:

PLEASE ENTER THE LAST TWO DIGITS OF YOUR HOME PHONE NUMBER AND THE LAST TWO DIGIST OF YOUR SOCIAL SECURITY NUMBER IN THE SPACES BELOW:

#____ ____ __ __ ____

TODAY'S DATE: _____   AGE (please fill in): _____

Date of Birth: _____Month _____Day _____Year

Gender: Female   Male

DIRECTIONS: Please check only ONE answer for each category.

Marital Status:

_____Single, (Never Married)
_____Single, (Divorced)
_____Married, never divorced
_____Married, prior marriage (s)
_____Widowed

ETHNIC BACKGROUND:

_____African American
_____American Indian or Alaskan native
_____Asian or Pacific Islander
_____Hispanic
_____Caucasian
_____Other: please specify:_____

LEVEL OF EDUCATION (Check highest level completed.)

_____Less than high school
_____High school
_____Some college
_____Associate degree
_____Baccalaureate degree
_____Master's degree
_____Doctoral degree

Has any family member ever had a depression or ever been treated for depression?

_____Yes   _____No   _____Don't Know

Have you EVER been evaluated, referred, or treated for depression?

_____Yes   _____No   PLEASE CONTINUE

Are you CURRENTLY being evaluated, referred, or treated for depression?

_____Yes   _____No

Are you CURRENTLY on medication to treat depression?

_____Yes   _____No

Are you CURRENTLY depressed?

_____Yes   _____No   _____Not Sure

ARE YOU TSU FACULTY or STAFF?
If YES, please complete below.
IF NO, CONTINUE TO PART II.

Please check current TSU position classification if you are employed either full-time or part-time at Towson State University:

_____Faculty

_____Professional Staff (Academic and/or Administrative support)

_____Staff

PLEASE CONTINUE

PART II

DIRECTIONS: PLEASE READ THE STATEMENTS CAREFULLY. CIRCLE THE DOT THAT BEST INDICATES HOW YOU FELT DURING THE PAST TWO DAYS. PLEASE SEE THE EXAMPLE BELOW. Please regard each line as representing the full range of each response.

EXAMPLE:

My favorite color is blue.

```
  .   (1)   2    3    4    5    6    7    8    9    10
  0                                                
Not at all                                Very Much
```

SECTION A.

1. I have sleep problems: trouble falling asleep/interrupted sleep/awaken too early.

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

2. I feel too tired to get through my usual day or need naps more often.

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

3. My body feels "heavy" (slowed down, sluggish, tired).

```
  0    1    2    3         5    6    7    8    9    10
Not at all                                Very Much
```

4. I am excited about doing the fun things (hobbies, interests, social contact) I usually do.

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

5. I have trouble with any activity (working/walking/talking) because it takes so much effort.

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

6. I feel restless (jittery/pace/wring my hands).

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

7. I think there is something wrong with my health.

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

8. I notice that everything seems gray/cloudy/drab/lacking color.

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

9. I notice attractive men/women as much as I used to.

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

10. I look forward to fun things (hobbies/interests/social contact) as much as I usually do.

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

11. I brood and worry a lot (think about the same thing over and over again).

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

12. I notice that food does not taste as good as usual.

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

13. I feel depressed (sad, blue, and gloomy).

```
  0    1    2    3    4    5    6    7    8    9    10
Not at all                                Very Much
```

14. I have trouble concentrating or remembering things more than usual.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

15. I have much to look forward to in the future.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

16. I have more personal faults and shortcomings than most people.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

17. I feel afraid for no reason.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

18. I am irritable (snap at people/pick arguments) more than usual.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

19. I feel like crying or have trouble controlling my crying.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

20. I feel responsible for bad things that have happened.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

SECTION B

21. I notice that I am losing weight without dieting.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

22. I have noticed that my dreams are stranger (disturbing/vivid/weird) or that I have more nightmares.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

23. I have increased cravings for sweets and carbohydrates.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

24. I have trouble getting out of bed and getting started in the morning.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

25. My appreciation of sounds (music/conversation/nature) is as good as ever.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

26. I get excited/happy when something good happens.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

27. I often have morbid thoughts (think about death or dying).

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

28. My interest in being sensual/physical/loving is worse than usual.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

29. I make decisions with difficulty.

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

30. My appetite has changed (either increased or decreased).

```
0   1   2   3   4   5   6   7   8   9   10
Not at all                      Very Much
```

PART III   SYMPTOMS

	NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS

DURING THE LAST WEEK HAVE YOU
EXPERIENCED....

		NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
A.	Sleep disturbance (trouble getting to sleep or staying asleep, early wakening, or sleeping longer than usual?	1	2	3	4
B.	Tired feeling (no energy)	1	2	3	4
C.	Slowed activity (sluggish movement or speech; hard to get moving) OR Hyperactivity (excessive movement/working/fidgeting)	1	2	3	4
D.	A change in appetite, weight, or the amount you eat (a decrease or increase that you did not plan)	1	2	3	4
E.	Constipation	1	2	3	4
F.	Headaches, or other aches and pains	1	2	3	4
G.	Muscle tension	1	2	3	4

		NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS

DURING THE LAST WEEK HOW OFTEN
HAVE YOU FELT...

		NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
A.	Sad (gloomy, discouraged, blue, numb, empty, or like you just don't care)	1	2	3	4
B.	That things aren't much fun anymore	1	2	3	4
C.	Down on yourself (worthless, low self-esteem, unlovable, self-disliking)	1	2	3	4
D.	Guilty (like a bad person)	1	2	3	4

	NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
DURING THE LAST WEEK HOW OFTEN HAVE YOU FELT...				
E. Unable to concentrate, think clearly, remember, or make decisions	1	2	3	4
F. Like crying (or that you would cry, but no longer seem able to)	1	2	3	4
G. Irritable (touchy, nervous, jittery)	1	2	3	4
H. Uninterested in people	1	2	3	4
I. Unattractive	1	2	3	4
J. Uninterested in sex	1	2	3	4
DURING THE LAST WEEK HAVE YOU FOUND YOURSELF...				
A. Thinking that something is wrong with you (you're inadequate—lacking something needed to be happy or successful)	1	2	3	4
B. Thinking that the world and its people are harsh and unfair	1	2	3	4
C. Thinking of suicide (either planning it, thinking you'd like to, or thinking it would be better if you weren't around)	1	2	3	4
D. Noticing mainly the negative aspects of situations or people	1	2	3	4
E. Worrying about your physical health or body functions	1	2	3	4

Please continue

PART IV

PLEASE CIRCLE THE ONE RESPONSE TO EACH ITEM THAT BEST DESCRIBES YOU FOR THE PAST SEVEN DAYS.

A. Falling Asleep:

0   I never take longer than 30 minutes
    to fall asleep.
1   I take at least 30 minutes to fall
    asleep, less than half the time.
2   I take at least 30 minutes to fall
    asleep, more than half the time.
3   I take more than 60 minutes to fall
    asleep, more than half the time.

B. Sleep During the Night:

0   I do not wake up at night.
1   I have a restless, light sleep with a
    few brief awakenings each night.
2   I wake up at least once a night, but
    I go back to sleep easily.
3   I awaken more than once a night and
    stay awake for 20 minutes or more,
    more than half the time.

C. Waking Up Too Early:

0   Most of the time, I awaken no more
    than 30 minutes before I need to get
    up.
1   More than half the time, I awaken
    more than 30 minutes before I need to
    get up.
2   I almost always awaken at least one
    hour or so before I need to, but I go
    back to sleep eventually.
3   I awaken at least one h our before I
    need to, and can't go back to sleep.

D. Sleeping Too Much:

0   I sleep no longer than 7-8
    hours/night, without napping during
    the day.
1   I sleep no longer than 10 hours in a
    24-hour period including naps.
2   I sleep no longer than 12 hours in a
    24-hour period including naps.
3   I sleep longer than 12 hours in a 24-
    hour period including naps.

E. Feeling Sad:

0   I do not feel sad.
1   I feel sad less than half the time.
2   I feel sad more than half the time.
3   I feel sad nearly all of the time.

F. Feeling Irritable:

0   I do not feel irritable.
1   I feel irritable less than half the
    time.
2   I feel irritable more than half the
    time.
3   I feel extremely irritable nearly all
    of the time.

G. Feeling Anxious or Tense:

0   I do not feel anxious or tense.
1   I feel anxious (tense) less than half
    the time.
2   I feel anxious (tense) more than half
    the time.
3   I feel extremely anxious (tense)
    nearly all of the time.

H. Response of Your Mood to Good or Desired
    Events:

0   My mood brightens to a normal level
    which lasts of several hours when
    good events occur.
1   My mood brightens but I do not feel
    like my normal self when good events
    occur.
2   My mood brightens only somewhat to a
    rather limited range of desired
    events.
3   My mood does not brighten at all,
    even when very good or desired events
    occur in my life.

I. Mood in Relation to the Time of Day:

0   There is no regular relationship
    between my mood and the time of day.
1   My mood often relates to the time of
    day because of environmental events
    (e.g., being alone, working).
2   In general, my mood is more related
    to the time of day than to
    environmental events.
3   My mood is clearly and predictably
    better or worse at a particular time
    each day.

J. The Quality of Your Mood:

0   The mood (internal feelings) that I
    experience is very much a normal
    mood.
1   My mood is sad, but this sadness is
    pretty much like the sad mood I would
    feel if someone close to me died or
    left.
2   My mood is sad, but this sadness has
    a rather different quality to it than
    the sadness I would feel if someone
    close to me died or left.
3   My mood is sad, but this sadness is
    different from the type of sadness
    associated with grief or loss.

Please complete either K or L (not both)

K. Decreased Appetite:

0   There is no change in my usual
    appetite.
1   I eat somewhat less often or lesser
    amounts of food than usual.
2   I eat much less than usual and only
    with personal effort.
3   I rarely eat within a 24-hour period,
    and only with extreme personal effort
    or when others persuade me to eat.

L. Increased Appetite:

0   There is no change from my usual
    appetite.
1   I feel a need to eat more frequently
    than usual.
2   I regularly eat more often and/or
    greater amounts of food than usual.
3   I feel driven to overeat both at
    mealtime and between meals.

Please continue

Please complete either M or N (not both)

M. Within the Last Two Weeks:

0   I have not had a change in my weight.
1   I feel as if I've had a slight weight loss.
2   I have lost 2 pounds or more.
3   I have lost 5 pounds or more.

N. Within the Last Two Weeks:

0   I have not had a change in my weight.
1   I feel as if I've had a slight weight gain.
2   I have gained 2 pounds or more
3   I have gained 5 pounds or more.

O. Concentration/Decision Making:

0   There is no change in my usual capacity to concentrate or make decisions.
1   I occasionally feel indecisive or find that my attention wanders.
2   Most of the time, I struggle to focus my attention or to make decisions.
3   I cannot concentrate well enough to read or cannot make even minor decisions.

P. View of Myself:

0   I see myself as equally worthwhile and deserving as other people.
1   I am more self-blaming than usual.
2   I largely believe that I cause problems for others.
3   I think almost constantly about major and minor defects in myself.

Q. View of My Future:

0   I have an optimistic view of my future.
1   I am occasionally pessimistic about my future, but for the most part I believe things will get better.
2   I'm pretty certain that my immediate future (1-2 months) does not hold much promise of good things for me.
3   I see no hope of anything good happening to me anytime in the future.

R. Thoughts of Death or Suicide:

0   I do not think of suicide or death.
1   I feel that life is empty or wonder if it's worth living.
2   I think of suicide or death several times a week for several minutes.
3   I think of suicide or death several times a day in some detail, or I have made specific plans for suicide or have actually tried to take my life.

S. General Interest:

0   There is no change from usual in how interested I am in other people or activities.
1   I notice that I am less interested in people or activities.
2   I find I have interest in only one or two of my formerly pursued activities.
3   I have virtually no interest in formerly pursued activities.

T. Energy Level:

0   There is no change in my usual level of energy.
1   I get tired more easily than usual.
2   I have to make a big effort to start or finish my usual daily activities (for example, shopping, homework, cooking or going to work).
3   I really cannot carry out most of my usual daily activities because I just don't have the energy.

U. Capacity for Pleasure or Enjoyment (including sex):

0   I enjoy pleasurable activities just as much as usual.
1   I do not feel my usual sense of enjoyment from pleasurable activities.
2   I rarely get a feeling of pleasure from any activity.
3   I am unable to get any pleasure or enjoyment from anything.

V. Interest in Sex (Please rate Interest, not Activity:

0   I'm just as interested in sex as usual.
1   My interest in sex is somewhat less than usual or I do not get the same pleasure from sex as I used to.
2   I have little desire for or rarely derive pleasure from sex.
3   I have absolutely no interest in or derive no pleasure from sex.

W. Feeling slowed down:

0   I think, speak, and move at my usual rate of speed.
1   I find that my thinking is slowed down or my voice sounds dull or flat.
2   It takes me several seconds to respond to most questions and I'm sure my thinking is slowed.
3.   I am often unable to respond to questions without extreme effort.

X. Feeling restless:

0   I do not feel restless.
1   I'm often fidgety, wring my hands, or need to shift how I am sitting.
2   I have impulses to move about and am quite restless.
3   At time, I am unable to stay seated and need to pace around.

Y. Aches and pains:

0   I don't have any feeling of heaviness in my arms or legs and don't have any aches or pains.
1   Sometimes I get headaches or pains in my stomach, back or joints but these pains are only sometime present and they don't stop me from doing what I need to do.
2   I have these sorts of pains most of the time.
3   These pains are so bad they force me to stop what I am doing.

Z.   Other bodily symptoms:

0   I don't have any of these symptoms: heart pounding fast, blurred vision, sweating, hot and cold flashes, chest, ringing in my ears, or shaking.

1   I have some of these symptoms but they are mild and are present only sometimes.

2   I have several of these symptoms and they bother me quite a bit.

3   I have several of these symptoms and when they occur I have to stop doing whatever I am doing.

AA.   Panic/Phobic symptoms:

0   I have no spells of panic or specific fears (phobia) (such as animals or heights).

1   I have mild panic episodes or fears that do not usually change my behavior or stop me from functioning.

2   I have significant panic episodes or fears that force me to change my behavior but do not stop me from functioning.

3   I have panic episodes at least once a week or severe fears that stop me from carrying on my daily activities.

BB.   Constipation/diarrhea:

0   There is no change in my usual bowel habits.

1   I have intermittent constipation or diarrhea which is mild.

2   I have diarrhea or constipation most of the time but it does not interfere with my day-to-day functioning.

3   I have constipation or diarrhea for which I take medicine or which interferes with my day-to-day activities.

CC.   Interpersonal Sensitivity:

0   I have not felt easily rejected, slighted, criticized or hurt by others at all.

1   I have occasionally felt rejected, slighted, criticized or hurt by others.

2   I have often felt rejected, slighted, criticized or hurt by others, but these feelings have had only slight effects on my relationships or work.

3   I have often felt rejected, slighted, criticized or hurt by others and these feelings have impaired my relationships and work.

DD.   Leaden Paralysis/Physical Energy:

0   I have not experienced the physical sensation of feeling weighted down and without physical energy.

1   I have occasionally experienced periods of feeling physically weighted down and without physical energy, but without a negative effect on work, school, or activity level.

2   I feel physically weighted down (without physical energy) more than half the time.

3   I feel physically weighted down (without physical energy) most of the time, several hours per day, several days per week.

Thank you.

Appendix C

Evaluation Studies of Self-Report Depression Scales

SCALE	REFERENCE	POPULATION AND SAMPLE	ANALYSES	RESULTS
Inventory to Diagnose Depression (IDD) (Zimmerman, Coryell, Corenthal and Wilson, 1986)	Zimmerman et al., 1986	Psychiatric inpatients diagnosed with a variety of psychiatric disorders (n=220) and normal controls (n=15)	Reliability 1) test-retest 2) split-half 3) Cronbach's alpha	$r = .98$ $r = .93$ $r = .92$
			Validity Correlation with other scales 1) Hamilton Rating Scale (HRS) 2) Beck Depression Inventory (BDI) 3) Carroll Rating Scale (CRS) 4) Sensitivity to change	Hamilton Rating $r = .80$ $r = .87$ $r = .81$ Significant correlations between scores on both the HRS and IDD from admission to discharge
IDD	Zimmerman and Coryell, 1987	Normals: non-patient, first-degree relative of schizo, psychotic, depressives and normal controls (n=39.8)	Reliability 1) internal consistency 2) item--total correlation Concurrent Validity	Split-half =.91 (Spearman-Brown) Cronbach's Alpha =.92 Median correlation coefficient =.47 97.2% rate of agreement between interview using Diagnostic Interview Scale (DIS) and IDD
Inventory for Depressive Symptomatology (IDS) (Rush, 1985)	Rush et al., 1986	Outpatient Clinic Patients of the Affective Disorder Unit, Dallas, TX (n=289) average age =38.2 yrs 64% female 14.3 yrs education 97.2% Caucasian 55% married	Reliability Internal Consistency 1) item-total correlation 2) Cronbach's Alpha Concurrent validity	 range of $r = .33 - .72$ $r = .85$

Appendix C   Evaluation Studies of Self-Report Depression Scales (continued)

SCALE	REFERENCE	POPULATION & SAMPLE	ANALYSES	RESULTS
		Normal controls (n=23) average age =40.4 65% female 15.2 yrs education 95.7% Caucasian 39.1% married	1) BDI and IDS 2) HRSD (observing, noting) and IDS Construct Validity 1) *t*-test 2) step-wise discriminant analysis 3) factor structure	Pearson r = .78 r = .67 significantly differentiated scores in depressed correctly classified different types of depression as well as BDI 1) mood 2) anxiety 3) endogenous symptoms 4) atypical features
Carroll Rating Scale (CRS) (Carroll, 1981)	Carroll, Feinberg, Smouse, Rawson and Greden (1981)	1) employees at the University of Michigan Medical Center, ages 18 to 64 (n =119) 2) patients being treated for depression (n =200)	Reliability internal consistency Validity correlation with Hamilton Rating Scale Factor Structure	Split-half coefficient = .87 r = .80 two interpreted factors *Factor 1:* Depression Severity *Factor 2:* Anxiety and Agitation
Center for Epidemiologic Studies Depression Scale (CES-D) (Radloff, 1977)	Radloff, 1977	Sample 1: Probability samples of households designed to be representative of Kansas City, Missouri and Washington County, Maryland. An individual aged 18 and over was selected randomly from each household.	Reliability 1) Coefficient alpha 2) test-retest	r = .85 general sample r = .90 patient sample r = .57 general sample r = .53 patient sample

SCALE	REFERENCE	POPULATION & SAMPLE	ANALYSES	RESULTS
		Sample 2: Psychiatric inpatients at Washington County (n =70) and New Haven, CN (n =35)	**Validity** 1) Concurrent validity between CES-D and Hamilton Rating Scale (observer rating) 2) Discriminant	$r = .44$ $r = .69$ to .75 at 4 weeks of treatment significantly differentiated depressed and non-depressed patients
CES-D	Breslau, 1985	Mothers of handicapped children aged 8 to 23 years. mean age =42 years 76% Caucasian 68% married mean years of school = 12 (n =310)	Cronbach's alpha Sensitivity--DSM-III criteria using DIS Specificity--structured interview	$r = .90$ 87.5% 73%
CES-D	Orme, Reis and Herz, 1986	Individual parents who participated in family support programs designed to prevent child abuse and neglect, located in an Illinois community with high rates of poverty. 89% female 46% Black 50% Caucasian mean age =21.3 years (n =116)	**Reliability** 1) Cronbach's alpha 2) Item--total correlation **Validity** 1) factor analyses 2) discriminant validity using multiple regression	$r = .88$ mean coefficient $r = .52$ items loaded on three factors: self esteem trait anxiety state anxiety CES-D was correlated at .71 with trait anxiety accounting for 49% of the variance.
CES-D	Coyle, 1990 (dissertation)	790 adults with a physical disability (aged 18 to 55 years)	**Validity** Factor structure using principal components analysis with onthogonal rotation. Based on Scree test factors named according to variables which loaded .40 or more on a factor. **Reliability** Cronbach's alpha	Found four factors 1) depressed 2) somatic 3) positive affect 4) interpersonal relationships $r = .90$

Appendix C   Evaluation Studies of Self-Report Depression Scales (continued)

SCALE	REFERENCE	POPULATION & SAMPLE	ANALYSES	RESULTS
Wakefield Self-Assessment Depression Inventory (1971) Description: a modification and shortening of Zung's Self-rating Depression Scale (SDS)	Snaith, Ahmed, Mehta and Hamilton, 1971	200 normal hospital employees (122 females; 78 males); 100 patients demographics not reported	1) Construct Validity 2) Test-retest of patients who had six ECT treatments 3) Convergent Validity	1) Differentiated between means of patients and normals 2) Correlation coefficient $r = .68$ 3) Correlation coefficient $r -.87$ with Hamilton Scale for Depression

Aitken's Visual Analogue Scale (VAS) (Aitken, 1969)
 	
Description: a straight line 100 mm in length with anchors depression absent at one end and extreme depression at other end.
	Little and McPhail, 1973	8 female outpatients diagnosed with depression	1) Correlation with three measures:
			    VAS by patient and psychiatrist
			    VAS with BDI
			    VAS by patient correlated with
			       VAS by psychiatrist
			2) Test-retest at monthly intervals
				$r = .80$
				$r = .76$
				
				$r = .76$
				No significant difference

Aitken's Visual Analogue Scale for Depression (VASD) Description: a 100 mm line with statement I have never felt more depressed and at the other I am not depressed now. Scale is scored by marking in millimeters	Davies, Burrows and Poynton, 1975	Depressed inpatients at university psychiatric unit (Australia). Patients middle/lower social class; 29 men, 43 women; mean age 37.6 (n=72)	1) Convergent Validity	Significant correlations (.51 - .88 range) with BDI and Zung SDS at 0, 7, 14, and 21 days correlated .188 with Hamilton Scale (interview) at 21 days.

Visual Analogue Mood Scale (VAMS)
 	
Description: a 100 mm line with anchors worst mood at left edge and best mood at right edge
	Luria, 1975	Psychiatric inpatients ages 20-70 at a New York hospital. Excluded alcoholics. Included non-affective psychiatric conditions (i.e., schizophrenia) (n=62)	1) Concurrent Validity correlate with SDS
			2) Test-Retest
			    2-hour test-retest
			    24-hour
				Correlation by VAMS and SDS: entire sample $r = .56$, affective psychoses $r = .63*$ (*higher score on SDS, the worse the mood) negative correlations expected (range) $r = .73-.91$ (range) $r = .56-.72$

Appendix C   Evaluation Studies of Self-Report Depression Scales (continued)

SCALE	REFERENCE	POPULATION & SAMPLE	ANALYSES	RESULTS
VAMS Description: used a rectangular card 100 mm by 35 mm on which the following instruction was printed: "How is your mood right now?" A mark on the line toward the left was <u>worst mood</u>; toward the right, <u>best mood</u>.	Folstein and Luria, 1975	**Sample A** 33 adult male patients in U.S. Armed Forces hospitalized on psychiatric or orthopedic wards of a naval hospital. Variety of psychiatric diagnoses in sample and those without psychiatric diagnosis. **Sample B** 31 patients hospitalized at New York hospital with major psychiatric disorders: affective psychoses, schizophrenia, neuroses and personality disorder.	<u>Validity</u> Correlated VAMS with the Self-rating Depression Scale (SDS) <u>Reliability</u> Test-Retest using product moment for within subject and within group for 15 days at 24-hour time intervals.	<u>Validity</u> Correlation coefficient: Sample A, $r = .64$; Sample B, $r = .67$ <u>Reliability</u> **Sample A:** $r = .61$ for within group $r = .32$ for within patient **Sample B :** $r = .73$ for within group $r = .48$ for within patient
Visual Analogue Scale for Depression (VASD) (Selth, 1990) Description: developed by author for a dissertation study. It is a 6.5 inch horizontal line representing the continuum of depressed affect, with endpoint labels of <u>not depressed</u> and <u>very depressed</u>.	Selth, 1990 (dissertation)	Elderly nursing home residents (n=59)	<u>Validity</u> Convergent correlations 1) VASD and Geriatric Depression Scale (GDS) 2) VASD and observer rating <u>Discriminant Validity</u> Analysis of variance	 $r = .68$ $r = .69$ Accurately discriminated subjects taking different antidepressant medication versus no psychotropic medication.

Appendix C   Evaluation Studies of Self-Report Depression Scales (continued)

SCALE	REFERENCE	POPULATION & SAMPLE	ANALYSES	RESULTS
Levine-Pilowsky Depression Questionnaire (1969) (Pilowsky, Levine & Boulton, 1969)	Pilowsky and Spaulding, 1972	Psychiatric inpatients and outpatients diagnosed with depression according to decision rule. Sample I n=80 (all females) mean age = 46.8 Sample II n=37 (13 males; 24 females)	Validity 1) t-test to discriminate 2) t-test to show sensitivity of depression scores to clinical progress 3) concurrent measure of observer ratings and depression scale scores	1) The depression group's scores were different from the non-depressed, but no significant difference was reported. 2) Significant changes in depression scores over time. Coefficient =.62
Self-rating Depression Scale (SDS) (Zung, 1965)	Zung, 1965	Inpatient hospitalized patients diagnosed with depression according to clinical diagnostic criteria based on factor analysis of depressive symptoms (n=56)	Validity Compared mean scores of patients with depression, patients with other psychiatric disorders, and normal controls using t-tests.	Significant differences among the three groups
SDS	Kozeny, 1987	Czech population which included normal and psychiatric inpatients diagnosed with depression at a psychiatric hospital (Prague) Sample I: Patients n=185 (32 males; 153 females) mean age = 33 years Sample II: Normals n=226 (54 males; 172 females) mean age = 28.1 years	Reliability Internal consistency 1) mean inter-item correlation 2) Cronbach's alpha Validity Factor analysis using direct oblimin rotation	Mean r =.18 with range of .22 to .71 r = .88 patients .82 normals Based on eigenvalues and tests of significance, found two factors: *Factor I*--psychomotor and psychological *Factor II*--affective and physiological

Appendix C   Evaluation Studies of Self-Report Depression Scales (continued)

SCALE	REFERENCE	POPULATION & SAMPLE	ANALYSES	RESULTS
SDS	Kivela and Pahkala, 1987	Elderly born 1923 or earlier, depressed according to DSM-III criteria. Living in Ahtari, a semi-industrialized community in Finland (n=1529)	**Construct Validity** Factor structure Principal - component factor analysis to determine eigenvalues greater than 1.0, orthogonal Varimax rotation. Algorithm using load in, at Gast 0.4000 for item inclusion.	Three factors emerged for both series: 1) depressed mood 2) loss of self-esteem and emptiness 3) irritability and agitation By sexes, four factors found for men: 1) depressed mood 2) fatigue and irritability 3) somatic symptoms 4) indecisiveness and hopelessness Four factors found for women: 1) loss of self-esteem,; hopelessness, and emptiness 2) depressed mood 3) fatigue 4) somatic Found different pattern of symptoms in men and women
SDS	Gabrys and Peters, 1985	Depressed, non-depressed and family members at a mental health center (Canada) Sample (n=587) 1) 218 non-depressed (115 females; 177 males); mean age =23.09 yrs 2) 369 depressed (192 females; 177 males); mean age =26.46 yrs 3) family escorts (173); mean age =44.59 yrs	**Reliability** 1) internal consistency 2) item-total correlations **Validity** predictive validity	r = family escorts .91 r = depressed clients .88 r = non-depressed .93 r = family escorts .80 r = depressed .82 r = non-depressed .85 Significant differences between non-depressed and depressed clients. In depressed clients, 8 % were false negatives. In non-depressed, 23 % scored above cut-off (false positive).

SCALE	REFERENCE	POPULATION & SAMPLE	ANALYSES	RESULTS
SDS	Holmes, Wurtz, Fouty and Burdick, 1988	n=671 women; mean age = 37.70 n=378 men; mean age = 38.86 Private psychiatric clinic in a mid-sized Midwestern city with a large state university Sample: 95% Caucasian, 4% Black, and 1% other race. Education: 21% < high school 29% high school diploma 23% some college 17% college 10% graduate degrees, including doctorates	Validity t-tests	Psychiatric patients scored significantly higher than nonpsychiatric patients.
SDS	Jonghe and Baneke, 1989	a psychiatric clinic (Amsterdam) Depressed and nondepressed patients (n=113); 85 depressed, 28 nondepressed; mean age = 32.8 yrs	Reliability Internal consistency Validity one-way ANOVA	1) Cronbach's alpha = .82 2) Split-half = .79 Significant differences between the depressed and nondepressed group. (F = 13.3, P = .0004) Overlap in group's scores

Appendix C   Evaluation Studies of Self-Report Depression Scales (continued)

SCALE	REFERENCE	POPULATION & SAMPLE	ANALYSES	RESULTS
SDS	Crittenden, Fugita, Bae, Lamug and Lin, 1992	University students in four countries: Korea, Philippines, Taiwan, and U.S. in intro level classes (n=966)	<u>Validity</u> ANOVA to compare SDS and subscale means across samples controlling for response set using ANCOVA <u>Reliability</u> Cronbach's alpha	After controlling for response set in U.S. and Taiwan, depression primarily reported in the psychological symptoms. Korea and Philippines = somatic first, followed by psychological symptoms. r = U.S. .84 r = Korea .82 r = Philippines .73 r = Taiwan .79 Conclusion: comparisons can be made with the SDS in different countries, but there is evidence that the symptoms vary with culture.
Beck Depression Inventory (BDI) (Beck, 1961)	Beck, Ward, Mendelson, Mock and Erbaugh, 1961	Psychiatric inpatients and outpatients of a university hospital (Pennsylvania), ages 15 - 44, lower SES, mostly Caucasian Variety of psych diagnoses Sample I = 226 Sample II = 183	1) internal consistency 2) correlation between BDI scores and clinician ratings of depression severity 3) assess changes in intensity of depression	All items significantly correlated positively with total score. Pearson r = .86 Spearman Brown 4 = .93 <u>Study I*</u> r = .65 <u>Study II**</u> r = .67 In 85% of cases, the change in depth of depression was predicted.

*   Except as noted, results are statistically significant at $p. < .05$ or lower.
**   Correlations reported as positive values unless specified otherwise.

Appendix C   Evaluation Studies of Self-Report Depression Scales (continued)

SCALE	REFERENCE	POPULATION & SAMPLE	ANALYSES	RESULTS
BDI	Hill, Kemp-Wheeler and Jones (1986)	1) College students at British university (n=160), males and females; ages 18-23 years 2) Patients (n=65), 44 women, mean age = 37.1; 21 men, mean age = 41.7 Diagnosed with depression and anxiety by a psychiatrist blind to the study.	Compared psychometrics on student and patient population. 1) factor analyses 2) used "Life Event Questionnaire" and "Eysenck Personality Inventory" to measure psychopathology in students and correlated with BDI.	1) Compared favorably for both students (7 factors) and patients (6 factors) 2) Found a correlation between psychopathology and BDI in students. Reasonable measure of depression in students, but use caution when using with college students.
BDI	Beck, Steer and Garbin, 1988	Psychiatric and nonpsychiatric subjects from the time period 1961 to June 1986 in research published studies with at least 30 patients.	1) Content Validity 2) Internal consistency 3) Concurrent Validity a) mean correlations between BDI and clinical ratings b) correlation between Hamilton Rating Scale for Depression (clinician) c) correlation with Zung SDS	1) Deliberately omitted questions on increased appetite and increased sleep since these can occur in normals. Agitation does not appear since it is inappropriate for self-report. 2) Mean coefficient Alpha $r = .86$ for psychiatric sample; $r = .81$ for nonpsychiatric sample. $r = .76$ psychiatric $r = .60$ nonpsychiatric $r = .73$ psychiatric $r = .80$ nonpsychiatric $r = .76$

Appendix C   Evaluation Studies of Self-Report Depression Scales (continued)

SCALE	REFERENCE	POPULATION & SAMPLE	ANALYSES	RESULTS
			4) Test-retest stability (varied from hours to weeks)	Ranged .48-.86 for psychiatric .60-.83 for nonpsychiatric
			5) Construct validity	Three factors extracted: negative attitude towards self, performance impaired, somatic disturbance.
			6) Discriminant validity (in a review of at least 10 research studies)	Distinguished psychiatric and nonpsychiatric patients and between different types of depression.
			7) Demographic correlates	Women, adolescents and Blacks scored higher on the BDI.

Appendix D

Panel of Reviewers

J. Raymond DePaulo, M.D.
Director, Affective Disorders Clinic
Professor, Department of Psychiatry
Johns Hopkins Medical Institutions
Baltimore, Maryland 21287

John Govern, Ph.D.
Assistant Professor
Department of Psychology
Towson State University
Towson, Maryland 21204

Neil Pauker, M.D.
Clinical Psychiatrist
7680 Cambridge Manor Place
Ft. Myers, Florida 33907

Deitra Wengert, Ph.D.
Assistant Professor
Department of Health Sciences
Towson State University
Towson, Maryland 21204

Dianne Taylor, RN, Ed.D
Assistant Professor
Department of Nursing
Towson State Univeristy
Towson, Maryland 21204

Sylvia Simpson, M.D.
Professor, Department of Psychiatry
Johns Hopkins Medical Institutions
Baltimore, Maryland 21287

Appendix E

Correa-Barrick Depression Scale (initial)

DIRECTIONS: PLEASE READ THE STATEMENTS CAREFULLY. CIRCLE
THE DOT THAT <u>BEST</u> INDICATES HOW YOU FELT DURING THE PAST <u>TWO</u>
<u>DAYS</u>. PLEASE SEE THE EXAMPLE BELOW.

EXAMPLE:
My favorite color is blue.
  Not at ALL _____ Very Much

TODAY'S DATE_____

1. I notice that I am losing weight without dieting.
    Not at All _____ Very Much

2. My appetite is as good as ever.
    Not at All _____ Very Much

3. I am eating more/have increased cravings for sweets,
    snacks/am gaining weight.
    Not at All _____ Very Much

4. I have strange (vivid/weird) dreams and disturbing
    nightmares.
    Not at All _____ Very Much

5. I have sleep problems: trouble falling asleep/
    interrupted sleep/awaken too early.
    Not at All _____ Very Much

6. I feel too tired to get through my usual day or need
    naps more often.
    Not at All _____ Very Much

7. My body feels "heavy" (slowed down, sluggish, tired).
    Not at All _____ Very Much

8. I am excited about doing the fun things (hobbies,
    interests, social contact) I usually do.
    Not at All _____ Very Much

9. I feel rested/pretty good in the morning.
    Not at All _____ Very Much

10. I have trouble with any activity (working/walking/
    talking) because it takes so much effort.
    Not at All _____ Very Much

11. I feel restless (jittery/pace/wring my hands).
    Not at All _____ Very Much

12. I think there is something wrong with my health.
Not at All _____ Very Much

15. I notice that everything seems gray/cloudy/drab/lacking color.
Not at All _____ Very Much

14. My sense of sound (music/conversation/nature) is as sharp/clear as ever.
Not at All _____ Very Much

16. I notice attractive men/women as much as I used to.
Not at All _____ _____ _____ Very Much

17. My ability to enjoy sex is worse than usual.
Not at All _____ Very Much

18. I look forward to fun things (hobbies/interests/social contact) as much as I usually do.
Not at All _____ Very Much

19. I brood and worry a lot (think about the same thing over and over again).
Not at All _____ Very Much

13. I notice that food does not taste as good as usual.
Not at All _____ Very Much

20. I still make decisions without any more difficultly than I usually have.
Not at All _____ Very Much

24. I feel depressed (sad, blue, and gloomy).
Not at All _____ Very Much

21. I have trouble concentrating or remembering things more than usual.
Not at All _____ Very Much

22. I have much to look forward to in the future.
Not at All _____ Very Much

23. I have more personal faults and shortcomings than most people.
Not at All _____ Very Much

26. My mood improves when something good happens.
Not at All _____ Very Much

27. I sometimes wish that I were dead.
Not at All _____ Very Much

28. I feel afraid for no reason.
    Not at All _____ Very Much

29. I am irritable (snap at people/pick arguments) more than
    usual.
    Not at All _____ Very Much

25. I feel like crying or have trouble controlling my
    crying.
    Not at All _____ Very Much

30. I feel responsible for bad things that have happened.
    Not at All _____ Very Much

Elsa Correa, M.D. & Christina Barrett Barrick, M.S., R.N.

PERMISSION TO COPY

I give permission to Christina B. Barrick to use and reproduce approximately 2,500 copies of the "Correa-Barrick Depression Scale" for the purpose of conducting validity and reliability studies for dissertation research. The copyright notice will appear on the scale.

Elsa Correa, M.D.

2/23/93
Date

Appendix F

Guidelines for Expert Reviewers

Thank you so much for agreeing to be a reviewer for the enclosed two self-rating depression scales, the "Depression Check-Up" and the "Correa-Barrick Depression Scale". The purpose of this study is to establish their validity and reliability and factor structure. Please consider the following in your review.

1. Face validity: Overall, does each scale appear to be measuring severity of depression?

2. Content validity:
    a. Does <u>each</u> scale meet the DSM-III-R criteria for major depression?
    b. Does <u>each</u> scale reflect qualitative aspects of patient experiences of depression?
    c. For <u>each</u> scale, should any item(s) be revised, deleted, or added? Please explain.

3. For the Correa-Barrick Scale <u>ONLY</u>:
    Do <u>each</u> of the items under "sensory/perceptual awareness" seem valid? (If uncertain, should each item be retained for statistical evaluation? Please explain).

4. For the Correa-Barrick Scale <u>ONLY</u>:
    The final scale will be down-sized to fewer items. Are there items that seem superfluous which could be tossed?

5. What clinical applications, if any, do you envision for <u>each</u> scale? (i.e. screening in primary care settings/assessment of patient responses to pharmacological/psychological therapies/adjunct to clinical interview for diagnosis, self-assessment, etc.?)

When you are finished, please send a letter of your analyses by <u>March 15</u> to me at the following address.

> 805 Chestnut Glen Garth
> Towson, Maryland 21204

Unless you disagree, your letter will appear in the appendices of the dissertation. Thank you for your time. I will be happy to share a brief report of the results and copies of the final scales if you are interested.

> Sincerely,
>
> Christina B. Barrick, MS, RN
> Ph.D. Candidate
> University of Maryland

Appendix G

Follow-up Letter to Each Panel Member

Guidelines for Expert Reviewers

current date

Thank you so much for agreeing to be a reviewer for the enclosed self-rating depression scale, "Correa-Barrick Depression Scale." The purpose of this study is to validate the scale. Please complete the following questions in the space provided and return in the enclosed SASE.

1. Face validity: Overall, does the scale appear to be measuring severity of depression?

2. Content validity: Do the scale items meet the DSM-III-R criteria for major depression?

3. Based upon patient report of experiences/symptoms, do each of the items listed under "sensory/perceptual awareness" seem valid?

4. Comments:

THANK-YOU

Christina B. Barrick, MS, RN
Ph.D. Candidate
University of Maryland

Diss. Advisor: Dr. Harvey Clearwater [telephone number].
Institutional Review Board [college/university; telephone number].

Appendix H

Letter regarding Beck Depression Inventory

THE PSYCHOLOGICAL CORPORATION®

555 ACADEMIC COURT, SAN ANTONIO, TEXAS 78204-2498
TELEPHONE: (512) 299-1061   TELEX: 5106015629 TPCSAT   FAX: (512) 270-0327

March 4, 1993

Ms. Christina B. Barrick
805 Chestnut Glen Garth
Towson, MD 21204

Dear Ms. Barrick:

Thank you for your February 27 follow-up letter concerning permission to use
the Beck Depression Inventory for testing purposes for use in your thesis
research.

In order to protect the combined usefulness of the test, and as a
responsible test publisher, we believe it is our responsibility to maintain
the security and integrity of our tests. Consequently, we cannot allow
items or portions of the test to be bound in, stapled with or microfilmed
with your thesis.

In addition, all testing should be conducted in your presence or that of
your faculty advisor so that all test materials remain in your hands.

Please be aware that we do not grant permission for reproduction of any of
our test materials. You will have to purchase all needed materials.

We will gladly grant permission for use of the test if the above
restrictions will be adhered to. Please sign and return a copy of this
letter to me for my files. You may then contact Sue Smith in Qualifications
at (800) 228-0752, ext. 293, to order your materials. You should request a
50% student discount from Mrs. Smith.

Also, please forward a copy of your thesis when it is completed so that I
may retain a copy in our library. If you have any questions regarding the
above please contact me directly.

Sincerely,

Christine Doebbler
Supervisor
Rights and Permissions

UNDERSTOOD AND AGREED

Christina B. Barrick   3/22/93
Name   Date

HARCOURT BRACE JOVANOVICH, INC.

Appendix I

Cover Letter, Sample II

current date

Dear Participant:

I am conducting a study on the validity and reliability of the enclosed survey which was developed by Dr. Correa and myself. I am asking for your help. Please complete the attached survey; it should take about 10 to 15 minutes for you to complete. Once you are finished, please seal the survey in the envelope and return to the office staff.

Your participation is completely voluntary and your refusal to participate will not affect your care in any way. You do not have to answer any question that you do not want to, and you may stop participating at any time.

If you have any questions or concerns abut the study, please feel free to contact me at [telephone number] or [name] Chairperson, Institutional Review Board, at [telephone number]. A written summary of the results of the study will be available upon request sometime in January 1994. At that time, please feel free to ask the office staff for a copy.

Your cooperation is greatly appreciated.

Thank you,

Christina B. Barrick, MS, RN
Assistant Professor

Appendix J

Summary Psychometric Evaluation for the **Depression Check-up** (Schiraldi, 1987), the **Correa-Barrick Depression Scale,** the **Correa-Barrick Depression Scale, Short Version**, the **Inventory for Depressive Symptomatology, Self-report** (Rush et al., 1986), and the **Beck Depression Inventory** (Beck et al., 1961)

Test	Depression Check-up			Correa-Barrick Depression Scale			Correa-Barrick Depression Scale, Short Version		Inventory for Depressive Symptomatology, Self-report		Beck Depression Inventory
	Pilot Sample (n=100)	Univ Sample (n=337)	Patient Sample (n=50)	Pilot Sample (n=100)	Univ Sample (n=337)	Patient Sample (n=50)	Univ Sample (n=337)	Patient Sample (n=50)	Univ Sample (n=337)	Patient Sample (n=50)	Pilot Sample (n=100)
Cronbach's alpha	.92	.94	.95	.92	.93	.96	.93	.95	.93	.91	.89
Test-retest reliability		.81*			.70*		.73*		.71*		
convergent validity with Inventory for Depressive Symptomatology, Self-report	.78	.88*	.85*	.72	.72*	.81*	.77*				
convergent validity with Beck Depression Inventory	.75*			.71*							
divergent validity: *t*-tests		*t* = 4.95*			NS		*t* = 3.00*		*t* = 5.95*		
factor analysis	three factors: (a) cognitive-emotional disturbance (b) psychophysiological symptoms (c) physiological symptoms			four factors: (a) cognitive-emotional disturbance (b) general outlook (c) physiological symptoms (d) sensory/perceptual disturbance							
depression item-total score correlation		.75*	.78*		.68*	.83*					

key: *   p < .01
NS   not significant

Appendix K

Letter regarding Depression Check-up

UNIVERSITY OF MARYLAND AT COLLEGE PARK

DEPARTMENT OF HEALTH EDUCATION

March 22, 1993

Christina Barrick
Dept. of Nursing
Towson State University
Baltimore, MD 21204

Dear Ms. Barrick,

I am pleased to grant you permission to duplicate and administer for research purposes only the Depression Check-up in the amount of 1500 copies. Please enter the following credit:

(c) 1990. Glenn R. Schiraldi, Ph.D., Dept. of Health Education, University of Maryland, College Park, MD 20742.
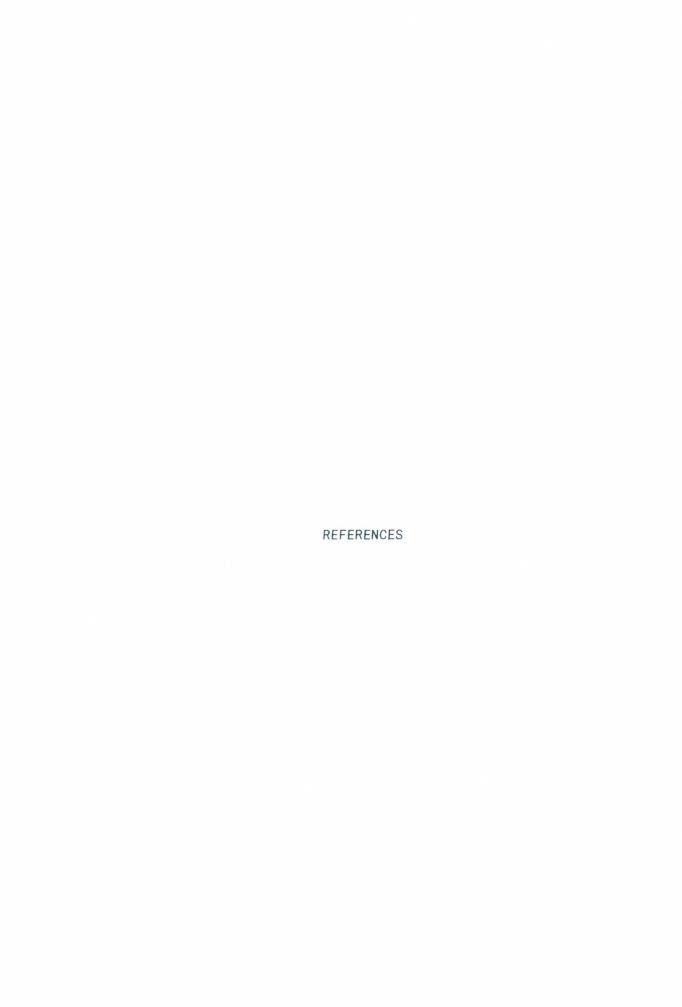
Please also note that the publisher of Hope and Help for Depression, Healthy People, Inc., is located in Miami Beach, FL.

Best wishes for your research.

Sincerely,

Glenn R. Schiraldi, Ph.D.
Faculty

REFERENCES

REFERENCES

Aitken, R. C. B. (1969). A growing edge of measurement of feelings using visual analogue scales. Proceedings of the Royal Society of Medicine, 62, 989--993.

Allen, R. J. (1983). Human stress: Its nature and control. Minneapolis: Burgess.

Alvarez, A. (1973). The savage god: A study of suicide. New York: Random House.

Ambelas, A., & George, M. (1988). Individualized stress vulnerabilities in manic-depressive patients with repeated episodes. Journal of the Royal Society of Medicine, 81, 448-449.

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington DC: Author.

American Psychological Association. (1985). Standards for educational and psychological testing. Washington DC: author.

Anastasi, A. (1988). Psychological testing. New York: Macmillan.

Angst, J., & Dobler-Mikola, A. (1984). The Zurich study: III--Diagnosis of depression. European Archives Psychiatric Neurological Science, 234, 30-37.

Authier, J., Bustafson, K., Guerney, B., & Kasdorf, J. (1975). The psychological practitioner as a teacher: A theoretical-historical and practical review. Counselling Psychologist, 5, 32-45.

Bandura, A., Taylor, C. B., Williams, S. L., Mefford, I. N., & Barchas, J. D. (1983). Catecholamine secretion as a function of perceived coping self-efficacy. Journal of Consulting and Clinical Psychology, 53(3), 406-414.

Basch, C. E. (1987). Focus group interview: An underutilized research technique for improving theory and practice in health education. Health Education Quarterly, 114(4), 411-448.

Bech, P., Kastrup, M., & Rafaelsen, D. J. (1986). Mini compendium of rating scales for states of anxiety, depression, mania, schizophrenia with corresponding DSM-III syndromes. Acta Psychiatrica Scandinavica, 73(Suppl. 326), 1--37.

Beck, A. T. (1976). Cognitive therapy and emotional disorders. New York: New American Library.

Beck, A. T., & Beamesderfer, A. (1974). Assessment of depression: The depression inventory. In P. Pichot (Ed.), Modern problems in pharmacopsychiatry. Basel, Switzerland: Karger.

Beck, A. T., & Steer, R. A. (1987). Beck depression inventory manual. San Antonio: Harcourt Brace Jovanovich.

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, 77-100.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. Archives of General Psychiatry, 4, 461--571.

Beck, K. H. (1989). Scale development and validation in health education research. Health Education Research, 4(3), 263-265.

Berne, R., & Levy, M. (1983). Physiology. St. Louis: Mosby.

Bidzinska, E. (1984). Stress factors in affective diseases. British Journal of Psychiatry, 144, 161-166.

Bond, A., & Lader, M. (1974). The use of analogue scales in rating subjective feelings. British Journal of Medical Psychology, 47, 211-218.

Breslau, N. (1985). Depressive symptoms, major depression, and generalized anxiety: A comparison of self-reports on CED-D and results from diagnostic interviews. Psychiatry Research, 15, 219-229.

Brink, T. L., Yesavage, J.A., Lum, O., Heersema, P.H., Adey, M. & Rose, T. L. (1982). Screening tests for geriatric depression. Clinical Gerontologist, 1, 37-43.

Cambell, D. T. (1963). Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), Psychology: A study of a science (pp. 94-172). New York: McGraw-Hill.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Carlsson, A. M. (1983). Assessment of chronic pain: I--Aspects of the reliability and validity of the visual analogue scale. Pain, 16, 87-101.

Carroll, B. J., Feinberg, M., Smouse, P. E., Rawson, S. G., & Greden, J. F. (1981). Carroll rating scale for depression: I--Development, reliability and validation. British Journal of Psychiatry, 138, 194-200.

Carroll, B. J., Fielding, J.M., & Blashki, T.G. (1973). Depression rating scales: A critical review. Archives of General Psychiatry, 28, 361-366.

Cella,. D. F., & Perry, S. W. (1986). Reliability and concurrent validity of three visual-analogue mood scales. Psychological Reports, 59, 827-833.

Clayton, P. (1986). Further comments: Psychosocial factors in a model of manic-depressive disease. Integrative Psychiatry, 4, 265.

Correa, E. (1983). Correa depression scale. Unpublished.

Coyle, C. P. (1991). Evaluating the psychometric properties of the Center for Epidemiological Studies: Depression scale for use with adults with physical disabilities. (Doctoral dissertation, Temple University, 1991). Dissertation Abstracts International, 52-02B, 1119.

Crittenden, D. S., Fugita, S. S., Bae, J., Lamug, C.B., & Lin, C. (1992). A cross-cultural study of self-report depressive symptoms among college students. Journal of Crosscultural Psychology, 23(2), 163-178.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Fort Worth: Holt, Rinehard & Winston.

Cronholm, B., & Daly, R. (1983). Evaluation of psychiatric treatment. In T. Helgason (Ed.), Methodology in evaluation of psychiatric treatment (pp. 3-32). Cambridge, England: Cambridge University Press.

Daley, D. C., Bowler, K., & Cahalane, H. (1992). Approaches to patient and family education with affective disorders. Patient Education and Counselling, 19, 163-174.

Davies, B., Burrows, G., & Poynton, C. (1975). A comparative study of four depression rating scales. Australian & New Zealand Journal of Psychiatry, 9, 21-24.

DeForge, B. R., & Sobal, J. (1988). Self-report depression scales in the elderly. Internatinoal Journal of Psychiatry in Medicine, 18(4), 325-338.

DePaulo, J. R., & Ablow, K. R. (1989). How to cope with depression. New York: McGraw-Hill.

DePue, R. A., & Evans, R. (1981). The psychobiology of depressive disorders: From pathophysiology to predisposition. Progress in Experimental Personality Research, 18, 1-114.

DeVellis, R. F. (1991). Scale development: Theory and application. CA: Sage.

Dixon, J. S., & Bird, H. A. (1981). Reproductibility along a 10 cm vertical visual analogue scale. Annals of the Rheumatic Diseases, 40, 876-89.

Dreyfuss, F., Dashber, H., & Assael, M. I. (1969). The relationship of myocardial infarction to depressive illness. Psychotherapy Psychosomatic, 17, 73-81.

Dunner, D., Patrick, V., & Fieve, R. (1979). Life events at the onset of bipolar affective illness. American Journal of Psychiatry, 136, 508-511.

Egeland, J. A., et al. (1987). Bipolar affective disorder linked to DNA markers on chromosome 11. Nature, 325, 783-787.

Ellis, A. (1973). Humanistic psychology: The rational-emotive approach. New York: Julian.

Endicott, J., Spitzer, R., Fleiss, J., & Cohen, J. (1976). The global assessment scale. Archives of General Psychiatry, 33, 766-771.

Everly, G. (1990). A clinical guide to the treatment of the human stress response. New York: Plenum Press.

Feighner, J. P., & Boyer W. F. (1991). Diagnosis of depression. New York: Wiley & Sons.

Fieve, R. (1975). Moodswing. New York: Bantam.

Folstein, M. F., & Luria, R. (1973). Reliability, validity, and clinical application of the visual analogue mood scale. Psychological Medicine, 3, 479-486.

Freyd, M. (1923). The graphic rating scale. Journal of Educational Psychology, 14, 83-102.

Friedman, A. S. (1963). Syndromes and themes of psychotic depression. Archives of General Psychiatry, 9, 504-509.

Gabrys, J. B., & Peters, K. (1985). Reliability, discriminant and predictive validity of the Zung self-rating depression scale. Psychological Reports, 57, 1091-1096.

Gift, A. G. (1989). Visual analogue scales: Measurement of subjective phenomena. Nursing Research, 38(5), 286-288.

Glassner, B., & Haldipur, C. V. (1983). Life events and late onset of bipolar disorder. American Journal of Psychiatry, 140(2), 215-217.

Goldberger, L., & Breznitz, S. (Eds.). (1982). Handbook of stress. New York: Macmillan.

Goodwin, F., & Jamison, K. (1990). Manic-depressive illness. New York: Oxford University Press.

Gove, W., McCorkel, J., Fain, T., & Hughes, M. (1976). Response bias in community surveys of mental health. Social Science and Medicine, 10, 497-502.

Green, L. W., & Lewis, F. M. (1986). Measurement and evaluation in health education and health promotion. Palo Alto, CA: Mayfield.

Grinker, R. R., Miller, J., Sabshin, M., Nunn, J., & Nunally, J. D. (1961). The phenomena of depression. New York: Harper.

Guyatt, G. H., Berman, L. B., Townsend, M., & Taylor, D. W. (1985). Journal Chronic Disease, 38(12), 1003-1007.

Guyatt, G. H., Townsend, M., Berman, L. B., & Keller, J. L. (1987). A comparison of likert and visual analogue scales for measuring change in function. Journal Chronic Disease, 40 (20), 1129-1133.

Hamilton, M. (1960). A rating scale for depression. Journal of Neurology, Neurosurgery & Psychiatry, 23, 56-62.

Hamilton, M. (1967). Development of a rating scale for pirmary depressive illness. British Journal of Social and Clinical Psychology, 6, 278-296.

Hasin, D., & Skodal, A. E. (1989). Standardized diagnostic interviews for psychiatric research. In C. Thompson (Ed.), The instruments of psychiatric research (pp. 19-57). Chicester: Wiley & Sons.

Hayes, M. H. S., & Patterson, D. G. (1921). Experimental development of the graphic rating method. Psychological Bulletin, 18, 98-99.

Hedlund, J. L., & Vieweg, B. W. (1979). The Zung Self-Rating Depression Scale: A comprehensive review. Journal of Operational Psychiatry, 10, 51-64.

Hill, A. B., Kemp-Wheller, S. M., & Jones, S. A. (1986). What does the Beck Depression Inventory measure in students? Personality Individual Differences, 7(1), 39-47.

Holmes, C. B., Fouty, H. E., Wurtz, P. J., & Burdick, B. M. (1988). Zung Self-Rating Depression Scale scores of psychiatric outpatients by age and sex. Psychological Reports, 62, 259-262.

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

Jonghe, J. F. M., & Baneke, J. J. (1989). The Zung Self-Rating Depression Scale: A replication study on reliability, validity and prediction. Psychological reports, 64, 833-834.

Joyce, C. R. B., Zutshi, D. W., Hrubes, V., & Mason, R. M. (1975). Comparison of fixed interval and visual analogue scales for rating chronic pain. European Journal of Clinical Pharmacology, 8, 415-420.

Kamenskaya, V. M., & Mikhailova, E. S. (Winter 1982). Study of the correlations among EEG and autonomic parameters in a stress situation in patients with various types of depression. Soviet Neurology & Psychiatry, 3-13.

Kaminsky. (1986). Psychotherapy in patients with endogenous depression. Unpublished manuscript, Johns Hopkins University, Department of Psychiatry and Behavioral Sciences, Baltimore.

Katon, W., & Sullivan, M. D. (1990). Depression and chronic medical illness. Journal of Clinical Psychiatry, 51 (Supplement 6), 3-11.

Kearns, N. P., Cruickshank, C. A., McGuigan, K. J., Riley, S. A., Shaw, S. P., & Snaith, R. P. (1982). A comparison of depression rating scales. British Journal of Psychiatry, 141, 45-49.

Keller, M. B., Klerman, G. L., Lavori, P. W., Fawcett, J. A., Coryell, W., & Endicott, J. Treatment received by depressed patients. Journal of the American Medical Association, 248, 1848-1855.

Keller, M. B., Lavori, P. W., Klerman, G. L. (1986). Low levels and lack of predictors of somatotherapy and psychotherapy received by depressed patients. Archives of General Psychiatry, 43, 458-466.

Kendall, P. C., Hollon, S. D., Beck, A. T., Hammen, C., & Ingram, R. E. (1987). Issues and recommendations regarding use of the Beck Depression Inventory. Cognitive Therapy and Research, 11(3), 289-299.

Kidder, L. H. (1981). Research methods in social relations. New York: Holt, Rinehart & Winston.

Kim, J., & Mueller, C. W. (1978). Introduction to factor analysis. CA: Sage.

Kivela, S., Pahkala, K. (1987). Factor structure of the Zung Self-Rating Depression Scale among a depressed elderly population. International Journal of Psychology, 22, 289-300.

Klerman, G. L., Weissman, M. M. (1989). Increasing rates of depression. Journal of the American Medical Association, 261, 2229-2235.

Kozeny, J. (1987). Psychometric properties of the Zung Self-Rating Depression Scale. Activitas Nervosa Superior (Praha), 29(4), 279-284.

Lambert, M. J., Christensen, E. R., & DeJulio, S. S. (1983). The assessment of psychotherapy outcome. New York: Wiley & Sons.

Lambert, M. J., Hatch, D. R., Kingston, M. D., & Edwards, B. C. (1986). Zung, Beck, and Hamilton rating scales as measures of treatment outcome: A meta-analytic comparison. Journal of Consulting and Clinical Psychology, 54(1), 54-59.

Lambert, J. J., Masters, K. S., & Astle, D. (1988). An effect-size comparison of the Beck, Zung, and Hamilton rating scales for depression: A three-week and twelve-week analysis. Psychological Reports, 63, 467-470.

Leserman, J., & Koch, G. (1993). Review of self-report depression and anxiety measures. Drug Information Journal, 27, 537-548.

Little, C. J., & McPhail, N. I. (1973). Measures of depressive mood at monthly intervals. British Journal Psychiatry, 122, 447-452.

Lopez-Ibor, J. J. (1972). Masked depression: 45th Maudsley Lecture. British Journal of Psychiatry, 120, 245-258.

Lubin, B. (1965). Adjective checklists for the measurement of depression. Archives of General Psychiatry, 12, 57-62.

Luria, R. E. (1975). The validity and reliability of the visual analogue mood scale. Journal of Psychiatric Research, 12, 51-57.

Magruder-Habib, K., Zung, W. W. K., & Feussner, J. R. (1990). Improving physicians' recognition and treatment of depression in general medical care. Medical Care, 28(3), 239-250.

Mausner, J. S., & Kramer, S. (1985). Epidemiology: An introductory text. Philadelphia: Saunders.

Mayer, J. (1978). Assessment of depression. In P. McReynolds (Ed.), Advances in psychological assessment (Vol. 4, pp. 298-425). San Francisco: Jossey-Bass.

McFarland, B. H., Freeborn, D. K., Mullooly, J. P., & Pope, C. R. (1989). The functioning and well-being of depressed patients: Results from the Medical Outcomes Study. Journal of the American Medical Association, 262, 914-919.

Mondimore, F. M. (1990). Depression: The mood disease. Baltimore: The Johns Hopkins University Press.

Montgomery, S., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. British Journal of Psychiatry, 137, 1081-1084.

Myers, J. K., Weissman, M. M., Tischler, G. L., Holzer, C. E., Leaf, P. J., Orvaschel, H., Anthony, J. C., Boyd, J. H., Burke, J. D., Kramer, M., & Stoltman, R. (1984). Six-month prevalence of psychiatric disorders in three communities. Archives of General Psychiatry, 41, 959-967.

Nielsen, A. C., Williams, T. (1980). Depression in ambulatory medical patients. Archives of General Psychiatry, 37, 999-1004.

Nunnally, J. C. (1967). Psychometric theory. New York: McGraw-Hill.

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

O'Connell, R. A. (1986). Psychosocial factors in a model of manic-depressive disease. Integrative Psychiatry, 4, 150-161.

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(1), 28-33.

Paykel, E. S. (1986). Further comments: Psychosocial factors in a model of manic-depressive disease. Integrative Psychiatry, 4, 264-265.

Payne, R. (1989). Glad to be yourself: A course of practical relaxation and health education talks. Physiotherapy, 75(10), 8-9.

Pilowsky, I., Levine, S., & Boulton, D. M. (1969). The classification of depression by numerical taxonomy. British Journal of Psychiatry, 114, 937-945.

Pilowsky, I., & Spalding, D. (1972). A method for measuring depression: Validity studies on a depression questionnaire. British Journal of Psychiatry, 121, 411-416.

Plutchik, R., & Praag, H. M. (1987). Interconvertability of five self-report measures of depression. Psychiatry Research, 22, 243-256.

Polit, D., & Hungler, B. (1983). Nursing research (2nd ed.). Philadelphia: Lippincott.

Post, R. D., Alford, C. E., Baker, N. J., Franks, R. D., House, R. M., Jackson, A. M., & Peterson, J. I. (1985). Comparison of self-reports and clinicians' rating of unipolar major depression. Psychological Reports, 57, 479-483.

Price, D. D., McGrath, P. A., Rafii, A., & Buckingham, B. (1983). The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain, 17, 45-56.

Rabkin, J., & Klein, D. (1987). The clinical measurement of depressive disorders. In A. Marsella, R. Hirschfeld, & M. Katz (Eds.), The Measurement of Depression (pp. 30-83). New York: Guldford Press.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. Applied Psychological Measurement, 1, 305-401.

Regier, D. A., Hirschfeld, R. M., Judd, L. L., Burke, J. D., Goodwin, F. K., & Lazar, J. B. (1988). The NIMH depression awareness, recognition, and treatment program: Structure, aims, and scientific basis. American Journal of Psychiatry, 145, 1351-1357.

Robins, L., Helzer, J., Croughan, J., & Ratcliff, K. (1981). National Institute of Mental Health diagnostic interview. Archives of General Psychiatry, 38, 381-389.

Roy, A., Guthrie, S., Pickar, D., & Linnoila, M. (1987). Plasma norepinephrine responses to cold challenge in depressed patients and normal controls. Psychiatry Research, 21, 161-168.

Rush, A. J. (March 25, 1993). Written reply from A. J. Rush to correspondence from C. Barrick.

Rush, A.J., Giles, D.E., Schlesser, M.A., Fulton, C..L., Weissenburger, J., & Burns, C. (1986). The inventory for depressive symptomatology (IDS): Preliminary findings. Psychiatry Research, 18, 65-87.

Schilddraut, J. (1965). The catecholamine hypothesis of affective disorders: A review of supporting evidence. American Journal of Psychiatry, 122, 509-522.

Schiraldi, G. (1987). Unpublished raw data on the Depression Check-up.

Schiraldi, G. (1990). Hope and help for depression: A practical guide. Miami Beach, FL: Healthy People.

Scott, J., & Hiskisson, E. C. (1979). Vertical or horizontal visual analogue scales? Annals of the Rheumatic Diseases, 38, 560.

Selth, C. (1990). Development of the behavioral observation of depression in the elderly scales (the Bodes) and validation of a visual analogue scale of depression with elderly nursing home residents (Doctoral dissertation, Pennsylvania State University, 1990). <u>Dissertation Abstracts International</u>, 5204B, 2313.

Shapiro, S., Skinner, E. A., Kessler, L. G., Von Korff, M., German, P. S., Tischler, G. L., Leaf, P. J., Benham, L., Cottler, L., & Regier, D. A. (1984). Utilization of health and mental health services: Three epidemiologic catchment area sites. <u>Archives of General Psychiatry</u>, <u>41</u>, 971-978.

Snaith, R. P., Ahmed, S. N., Mehta S., & Hamilton, M. (1971). Assessment of the severity of primary depressive illness. <u>Psychological Medicine</u>, <u>1</u>, 143-149.

Southwick, S., Mason, J. W., Giller, E. L., & Kosten, T. R. (1989). <u>Biological Psychiatry</u>, <u>25</u>, 67-74.

Spielberger, C. D. (Ed.). (1972). <u>Anxiety: Current trends in theory and research</u> (Vol. 1). New York: Academic.

Sriwatanakul, K., Kelvie, W., Lasagna, L., Calimlim, J., Weis, G., & Mehta, G. (1983). Studies with different types of visual analog scales for measurement of pain. <u>Clinical Pharmacology and Therapeutics</u>, <u>34</u>(2), 234-239.

Steer, R. A., & Beck, A. T. (1985). Modifying the Beck Depression Inventory: Reply to Vredenburg, Krames, and Flett. <u>Psychological Reports</u>, <u>57</u>, 625-626.

Steer, R. A., Beck, A. T., Riskind, J. H., & Brown, G. (1986). Differentiation of depressive disorders from generalized anxiety by the Beck Depression Inventory. <u>Journal of Clinical Psychology</u>, <u>42</u>(3), 475-478.

Teuting, P., & Koslow, S. (1983). <u>Special report on depression research</u> (Science Reports, National Institute of Mental Health; DHHS Publication No. 83-1085). Washington DC: U.S. Government Printing Office.

Thompson, C. (1989). <u>The Instruments of psychiatric research</u>. Chichester: Wiley & Sons.

Tsuang, M. T., Woolson, R. F., & Fleming, J. A. (1980). Premature deaths in schizophrenia and affective disorders: An analysis of survival curves and variables affecting a shortened survival. <u>Archives of General Psychiatry</u>, <u>37</u>, 979-983.

Weeke, A., Juel, K., & Vaeth, M. (1987). Cardiovascular death and manic-depressive psychosis. Journal of Affective Disorders, 13, 287-292.

Weissman, M. M., Leaf, P. J., Tischler, G. L., Blazer, D. G., Karno, M., Bruce, M. L., & Florio, L. P. (1988). Affective disorders in five United States communities. Psychological Medicine, 18, 141-153.

Wells, K. B., Stewart, A., Hays, P. D., Burnam, M. A., Rogers, W., Daniels, M., Berry, S., Greenfield, S., & Ware, J. (1989). The functioning and well-being of depressed patients: Results from the Medical Outcomes Study. Journal American Medical Association, 252(7), 914-919..

Wilson, H. S. (1985). Research in nursing. CA: Addison-Wesley.

Wing, J. K. (1970). A standard form of psychiatric present state examination. In E. H. Hare & J. K. Wing (Eds.), Psychiatric epidemiology, pp. 93-131. London: Oxford University Press.

Zealley, A. K., & Aitken, R. C. B. (1969). Measurement of mood. Proceedings of the Royal Society Medicine, 62, 993-996.

Zimbardo, P. G. (1988). Psychology and life. Glenview, IL: Scott, Foresman.

Zimmerman, M., & Coryell, W. (1987). The inventory to diagnose depression (IDD): A self-report scale to diagnose major depressive disorder. Journal of Consulting and Clinical Psychology, 55(1), 55-59.

Zimmerman, M., Coryell, W., Corenthal, C., & Wilson, S. (1986). A self-report scale to diagnose major depressive disorder. Archives of General Psychiatry, 43, 1076-1081.

Zung, W. W. (1968). Evaluating treatment methods for depressive disorders. American Journal of Psychiatry, 124, 40-47.

Zung, W. W. (1972). The depression status inventory: An adjunct to the self-rating depression scale. Journal of Clinical Psychology, 28, 539-543.

Zung, W. W. K. (1965). A self rating depression scale. Archives of General Psychiatry, 12, 63-70.

Zung, W. W. K. (1967). Factors influencing the self-rating depression scale. Archives of General Psychiatry, 16, 543-547.

Zung, W. W. K. (1990). The role of rating scales in the identification and management of the depressed patient in the primary care setting. <u>Journal of Clinical Psychiatry</u>, <u>51</u>(suppl 6), 72-76.

Zung, W. W. K., Magill, M., Moore, J., Moore, J. T., & George, D. T. (1983). Recognition and treatment of depression in a family medicine practice. <u>Journal of Clinical Psychiatry</u>, <u>44</u>, 3-6.

Zung, W. W. K., & Zung, E. M. (1986). Use of the Zung self-rating scale depression scale in the elderly. <u>Clinical Gerontologist</u>, <u>5</u>(1/2), 137-148.