

1991

# The Validity of Studies that Test the Effectiveness of Psychotherapy

Ian K. Isaacs

*Eastern Illinois University*

This research is a product of the graduate program in [Psychology](#) at Eastern Illinois University. [Find out more](#) about the program.

---

## Recommended Citation

Isaacs, Ian K., "The Validity of Studies that Test the Effectiveness of Psychotherapy" (1991). *Masters Theses*. 2254.  
<https://thekeep.eiu.edu/theses/2254>

This is brought to you for free and open access by the Student Theses & Publications at The Keep. It has been accepted for inclusion in Masters Theses by an authorized administrator of The Keep. For more information, please contact [tabruns@eiu.edu](mailto:tabruns@eiu.edu).

THESIS REPRODUCTION CERTIFICATE

TQ: Graduate Degree Candidates who have written formal theses.

SUBJECT: Permission to reproduce theses.

The University Library is receiving a number of requests from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow theses to be copied.

Please sign one of the following statements:

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

7/26/91  
Date

  
Author

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_  
Date

\_\_\_\_\_  
Author

The Validity of Studies that Test the

Effectiveness of Psychotherapy

(TITLE)

BY

Ian K. Isaacs

**THESIS**

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF

Master of Arts

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY  
CHARLESTON, ILLINOIS

1991

YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING  
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

7/22/91

DATE

2/26/91

DATE

ADVISER

DEPARTMENT HEAD

Validity

The validity of studies that test the  
effectiveness of psychotherapy

Ian K. Isaacs

Eastern Illinois University

Running head: Validity

## Abstract

This thesis is a review of the validity of studies on the effectiveness of psychotherapy. Ten randomly selected studies are examined to determine if the validity of the study was affected by threats to statistical conclusion, internal, construct, and external validity. All of the studies examined were published after 1985. The results indicate that validity of the research on the effectiveness of psychotherapy is affected by some threats to validity: History, Testing, and Selection bias, (ie. internal validity). In addition, external validity was affected by the lack of use of appropriate control group, lack of a description of the therapy used and the condition under which it was used, and a lack of a description of placebo group used. The evidence from this thesis suggests a need for improvement in the design of efficacy studies. In addition, the evidence suggests the need for better data reporting so that the research can be replicated.

### Acknowledgement

I respect and appreciate the work of all the researchers whom, through their study, contributed to this thesis. Special thanks to Dr. Charles Blaich who encouraged me to be more attentive to the literature on psychotherapy research. I would also like to thank Dr. Blaich for his direction, for without his guidance, this thesis may not have been completed.

Thank you Dr. Charles Blaich, Dr. Richard Cross, and Dr. James Havey for agreeing to be a part of my thesis committee, your input was invaluable to this thesis. I would like to thank all the professors whom I conferred with, your input was greatly appreciated.

I would also like to thank Dr. Fred Yaffe, Chair of the psychology department, and all other members of the psychology department at Eastern Illinois University for the opportunity to do this thesis.

Table of Contents

	Page
Chapter I Introduction	
Background	
Information.....	3
Description and Importance of the types of Validity.....	4
Threats to Validity.....	6
Chapter II Method	
Studies.....	16
Procedure.....	16
Criterion.....	17
Chapter III Results	
Findings.....	18
Chapter IV Discussion	
Implications of the thesis.....	28
Limitations.....	38
Suggestions for further Research.....	39
Reference.....	40
Appendix 1.....	43
Appendix 2.....	48

## Chapter I

## Introduction

## Background information

Studies on the general effectiveness of psychotherapy originated with Eysenck's (1952) highly publicized study. Eysenck argued that published research up to the time, most of which was on psychoanalytic therapy, was unable to demonstrate unequivocally that psychoanalytic psychotherapy worked. Since Eysenck's study, thousands of other studies have been conducted on this topic. For example, in their meta-analysis on the effectiveness of psychotherapy, Smith and Glass, (1977) identified 1,000 studies, and Shapiro and Shapiro, (1982) identified 400. The majority of subsequent studies have come to different conclusions than Eysenck. This later research, primarily meta-analyses, focused on the magnitude of the effectiveness of psychotherapy based on the findings of several studies. The researchers of these meta-analyses however, did not examine the validity of studies used in their meta-analyses to see that general conditions of validity were met.

According to Cook and Campbell, (1979) "decisions about whether a presumed cause and effect covary logically precede decisions about how strongly they covary." p. 41. A careful examination of the validity of studies that measure



psychotherapeutic effectiveness is important because it verifies that variations in patients' condition covary with variations in treatment. This, gives validity to claims that measured effects are principally due to the agency of the treatment rather than extraneous and uncontrolled influences. A randomly selected, representative sample, of the present research on the effectiveness of psychotherapy was examined to verify that basic areas of validity were met. The results indicated that fundamental areas of validity were affected by some of the threats to validity addressed in this thesis. The evidence from this thesis suggest that researchers need to better control for some threats to validity, and report more complete data so that their studies can be replicated to verify the findings.

#### Description and importance of the types of validity

Cook and Campbell (1979) discuss the four types of validity: Statistical Conclusion, Internal, Construct, and External validity. In terms of psychotherapy research, these types of validity address the following issues:

##### 1. STATISTICAL CONCLUSION VALIDITY

Statistical conclusion validity addresses the issue of covariation. It addresses the sensitivity of the

dependent variable to correctly measure variations in the independent variable, evidence to support presumed cause and effect between the dependent and the independent variable, and the strength of the covariation between the dependent variable and the independent variable.

## 2. INTERNAL VALIDITY

Internal validity deals with causation, specifically, the direction of causal relationship between treatment and subjects' condition. In psychotherapy research it deals with whether increases in treatment cause greater improvement subjects' condition.

## 3. CONSTRUCT VALIDITY

Construct validity addresses the issue of whether effects thought to be caused by one construct (eg. therapy) can be interpreted as being caused by other constructs (eg. placebo).

## 4. EXTERNAL VALIDITY

External validity addresses the ability to generalize findings across subjects, settings, populations, etc.

All four types of validity are relevant to this thesis because they address the following questions:

1. How reliability is the instrument used to measure variations in the independent variable? (ie. statistical conclusion validity).
2. Did variations in treatment covary with variations in the patient's condition? (ie. internal validity)
3. If variations in treatment covary with variations in the patient's condition, can we conclude that it is a therapeutic effect? (ie. construct validity).
4. Can we generalize the findings about treatment across settings and populations? (ie. external validity).

### Threats to validity

This thesis will examine some of the threats to the four types of validity. Kirk, (1982) lists threats to the four types of validity. All of the threats listed by Kirk, (1982) are not relevant to this thesis. However, the following threats are relevant as they address the questions raised at the end of the previous section.

#### 1. STATISTICAL CONCLUSION VALIDITY

##### A. The Reliability of the Measure

One threat to Statistical conclusion validity is low reliability of the measure. Reliability of the measure is when one cannot rely upon the

dependent variable to correctly measure changes in the subject's condition. Suppose a researcher tests the ability of passive disorder patients to identify assertive statements on an assertiveness test after therapy. The researcher tests the patients weekly on a test that is not a reliable test. It is possible that the test may measure other variables. In addition to measuring one's ability to identify assertive statements.

Furthermore, it may not be possible to know when the test is measuring the patient's ability to identify assertive statements and when it is measuring some other variable. Thus, researchers would not be sure that variations in the patient's condition were entirely due to variations in treatment. Since the reliability of standardized tests are usually verified, studies are judged as meeting this criterion if the researcher used a standardized test, or if the reliability of the instrument/test used to measure the effectiveness of therapy is verified.

## 2. INTERNAL VALIDITY

### A. History

History is a threat when there is a possibility

that outside events occurring after the beginning of treatment affect the dependent variable. Suppose a patient, who was successfully treated for depression after his first wife left him, is being treated for depression again, because his second wife left him. The patient received 6 months of therapy with little success. The patient receives a call from his previous therapist, who informs him that he must practice the techniques that he learned in his past treatment for therapy to work. The therapist also informs that patient, that his first wife promises to take him back if he recovers from his depression. The patient practices the techniques he learned from his previous therapist, and one week later the patient fully recovers from his depression. The patient's present therapist, unaware of what had transpired, attributes his recovery to therapy. Studies are judged as meeting this criterion if the researchers use patients who are not previously treated for the same or similar condition tested by the researcher.

#### B. Maturation

Maturation is a threat when changes in the subject, such as growing older, stronger, larger,

etc., affect the dependent variable. Suppose a well-educated student, who as a junior in high school takes the SAT exam, scores low because he is unable to handle the stress of taking the exam. The same student, now a year older, more mature and confident in his ability, takes the test his senior year and scores higher. All of the studies examined in this thesis measure the effect of psychotherapy on a short term basis, (less than one year). Recognizing that maturation is always occurring, this criterion is set very low to avoid making it impossible for any study to meet the criterion. Studies are judged as meeting this criterion if the researchers do not allow more time to elapse before taking the final measure than they allowed to elapse between treatment applications.

#### C. Multiple Testing

Multiple testing is a threat when subjects are repeatedly tested on the same test, and subjects become familiar with the test, and alter their response to correctly respond to test questions. Suppose the army reports that 90% of its cadets run an obstacle course in 32.5 seconds. However, the

cadets are tested on the obstacle course two-times-weekly for eight weeks before they are timed. The eight-week-period gives cadets time to learn the obstacle course which gives them an advantage over other individuals with less experience at running the same course. Studies are judged to meet this criterion if patients are not tested more than twice (pre-treatment and post-treatment) on the test used to take the final measure of the effectiveness of therapy.

D. Instrumentation

Instrumentation is a threat when the dependent variable is affected by, "(1) changes in the calibration of measuring instruments, (2) shifts in the criteria used by observers or scores, (3) and shifts in the metric in different ranges of a test." (Kirk, 1982). As an example of (1), suppose a therapist, treating a client for bulimia, weighs the client three times each week using the same scale. Suppose the scale breaks. Now, instead of the scale reading zero when no one is on it, the scale now reads 5 lbs. The therapist, unaware that the scale is broken, thinks that the client is improving because she has gained 5 lbs. As an

example of (2), suppose a doctor, practicing medicine in the 60's, admitted all patients with a body temperature greater than 99 degrees. He did so because of the inability of the medical community, at that time, to ward off serious illness due to a lack of knowledge of all deadly disease that existed. As a result of admitting these patients, the doctor saw over 3,000 patients per year. Later in his career, the doctor began to admit only those patients with a body temperature greater than 105 degrees because he was more comfortable with his knowledge and ability to treat most illness. As a result, the doctor only saw about 1500 patients per year. The doctor, unaware of his change in criteria for admittance, summarized in a report to a medical journal that the number of patients admitted to hospitals because of abnormally high body temperature had decreased over the past 20 to 30 years. As an example of (3), suppose a doctor, testing the effectiveness of drug X to improve an athlete's ability to run the mile, administers the drug to athletes and records the amount of time it takes the athletes to run the mile. The doctor records the amount of time it takes the athletes



to run a mile on a track measured in yards, and finds that the athletes run the mile in 4 minutes and 40 seconds. The doctor then records the amount of time it takes the athletes to run the mile on a track measured in meters. Because a metric mile is shorter, the doctor finds that athletes run the mile in 4 minutes and 20 seconds. The doctor markets drug X as a drug that has proven to be effective at increasing an athlete's ability to run the mile faster. Studies are judged as meeting this criterion if all patients are tested using the same procedure, and if the criteria of what constitutes a therapeutic effect is the same for all patients, based on the data reported by the researchers.

#### E. Selection Bias

Selection bias is a threat when patients in the treatment group are different from patients in the control group. Suppose a study is conducted to test the accuracy of college entrance exams to predict success in college. A researcher, testing high school seniors, places students with 4.0 G.P.A.'s in the treatment group, and places students with 2.0 G.P.A.'s in the control group. The researcher finds that students in the treatment

group, score higher on the exam than students in the control group. The researchers conduct a follow-up study after two years of college and find that students who were in the treatment group are doing much better in college than students who were in the control group. The researchers conclude that college entrance exams are good predictors of a student's success in college. Studies are judged as meeting this criterion if patients in all treatment conditions have the same symptoms and/or are not selected from different populations.

#### D. Randomization

Randomization is a threat when subjects are not randomly assigned to treatment conditions. Suppose a researcher conducts a study of all college students on their knowledge of Afro-American history. The researcher assigns one hundred students, in alphabetical order, by last names, to two groups. It so happens that 40 of the 50 students in the experimental group are Afro-American majors. The researchers conclude that college students are very knowledgeable of Afro-American history. Studies are judged as meeting this criterion if patients are randomly assigned to the

treatment conditions.

### 3. CONSTRUCT VALIDITY

#### A. Mono-Operation Bias

Mono-operation bias is a threat when researchers use only one treatment condition (ie. therapy) to account for variations in patients' conditions. It is important that researchers use multiple treatment conditions to account for variations in patients' conditions to avoid erroneously attributing effects caused say by, uncontrolled variables, to the treatment tested. By using multiple treatment conditions to account for the measured change in patients' conditions, researchers are able to verify that treatment condition A is what is causing improvement in the patients' condition than treatment condition B or C. Studies are judged as meeting this criterion if the researchers use more than one treatment condition, (ie. a control and/or a placebo group).

### 4. EXTERNAL VALIDITY

#### A. Interaction of Selection and Treatment

Interaction of selection and treatment is a threat when results obtained, using subjects with some specific characteristic, (ie. race, age, sex,

etc.), are erroneously generalized to other groups with different characteristics. Suppose after an 8 week reading course, 1,000 18-year-old Spanish-American males are tested to see how fast they can read. The results indicate that these individuals read approximately 75 words per minute. The researchers conclude that all American males read approximately 75 words per minute. Studies are judged as meeting this criterion if the researchers express caution when generalizing their findings to a different population than is used in the study.

B. Interaction of Setting and treatment

Interaction of setting and treatment is a threat when findings obtained in one setting are erroneously generalized to other settings. Suppose the government tests a new substance abuse treatment technique at a residential substance abuse clinic. The clinic's residents are individuals who have been placed there by court order for substance abuse. Ninety percent of the individuals do not believe that they need treatment. The technique does not prove to be effective. Because of this the government concludes that it will not provide funds to facilities that use this technique.

Studies are judged as meeting this criterion if the findings are not erroneously generalized to settings that are different from the setting of the study.

Two additional criteria are used, those being the mode of therapy used by the researcher, and the level of pathology of the subjects. These criteria are important in psychotherapy research because it is easier to replicate and check the validity of the findings when a standard mode of psychotherapy is used. By using subjects that meet a specific criterion, (ie. DSM-III) for a given disorder, the findings can be generalized to a specific population of patients who meet the criterion for that disorder. Studies meet the criterion of using a standard mode of therapy if researchers use a common mode of therapy (e.g. Cognitive, Behavioral, Psychoanalytic therapy). Studies meet the criterion of using patients with appropriate levels of pathology if patients' levels of pathology are measured on a standardized test, (ie. DSM-III).

## Chapter II

## Method

Studies

Ten studies were randomly selected for review, (See Appendix 1). The 10 studies examined the effectiveness of psychotherapy for the treatment of the following disorders: Four studies examined bulimia, one study examined agoraphobia, one study examined sociopathy, two studies examined chronic pain, and two studies examined depression. Seven of the ten studies were published in 1985, one study was published in 1988, and two studies were published in 1990. Four studies were published in psychiatric journals (ie. Archives of General Psychiatry), and six studies were published in psychological journals (ie. Journal of Consulting and Clinical Psychology).

Procedure

Ninety-three research articles on the effectiveness of psychotherapy, were obtained by accessing two data-base services. The two data-base services were INDEX MEDICUS and MEDLINE. INDEX MEDICUS is the National Library of Medicine's monthly bibliography of the literature of biomedicine. MEDLINE is an international data-base containing some 3,500 journals. MEDLINE service is a part of INDEX MEDICUS, International Denture Literature, and

International Nursing Index. Psychological Abstracts were also accessed, however, no research articles were found. The articles found in Psychological Abstracts were all review articles and Meta-analyses. The studies were obtained under the search title of "Psychotherapy Outcome Studies." All studies published before 1985 were omitted. Studies published before 1985 were omitted to avoid reviewing studies that had already been reviewed in meta-analyses. Studies that were classified as review articles, (ie. studies that were not research studies), were also omitted. Of the ninety-three articles, 24 articles were randomly selected using a random number generating system. From those 24 articles, 10 were randomly selected for examination.

The 10 studies selected included 5 studies that examined the effectiveness of psychotherapy for the treatment of Bulimia Nervosa. To get a more evenly distributed sample of the different types of studies, one of the 5 studies on bulimia was randomly omitted. Another study was randomly selected from the remaining studies on the original list.

#### Criteria

The validity of the studies was evaluated on the threats to validity discussed in the introduction. See Appendix 2 for critique of the studies on the threats to validity

examined.



## Chapter III

## Results

Table 1 presents the percentages of studies that may or may not have been affected by threats to validity. "Yes" indicates that the researchers adequately controlled for the threat. "No" indicates that the validity of the study may have been affected by the threat. "NC" indicates that it was not clear as to whether the validity of the study was affected by the threat. Table 2 presents studies that met each criterion. Seven of the eighteen threats were adequately controlled for in all ten studies: maturation, testing, instrumentation, randomization, interaction of selection and treatment, and interaction of setting and treatment. The results indicate that some of the studies did not meet the criteria for statistical, internal, construct, and external validity.

Statistical conclusion validity

The dependent measure used by 20 percent of the studies may not have been reliable. For example, Dedman, Numa, and Wakeling (1988) examined cognitive behavioral treatment for Bulimia Nervosa. The researchers' primary measure of the effectiveness of treatment was based on self-reports by the patients, with reference to their daily diary. Self-reported data are not reliable. Patients may report

information that makes them look good to the researcher and vice versa, depending on their view as to whether they need therapy or not, their view of the effectiveness of therapy, etc. Moore and Chaney, (1985) examined cognitive-behavioral therapy for the treatment of chronic pain. The researchers' primary measure to the effectiveness of treatment was also based on self reported data.

Table 1

The percentage of studies that met each criterion

Criterion	Yes	No	Not
-----------	-----	----	-----

clear

Statistical conclusion Validity:

Reliability of the Measure	80	20	
----------------------------	----	----	--

Internal Validity

History	50	50	
---------	----	----	--

Maturation	100	0	
------------	-----	---	--

Testing

1.	50	50	
----	----	----	--

2.	100	0	
----	-----	---	--

Instrumentation

1.	100	0	
----	-----	---	--

2.	90	10	
----	----	----	--

Selection Bias

1.	90	10	
----	----	----	--

2.	60	30	10
----	----	----	----

Randomization	100	0	
---------------	-----	---	--

Construct Validity

Use of placebo and/or control group	30	70	
-------------------------------------	----	----	--

External Validity

Interaction of Selection and Treatment

1.	100	0	
----	-----	---	--

	Validity		23
2.	100	0	
Interaction of Setting and Treatment	100	0	
Additional Criterion			
Mode of Therapy Used	90	0	10
Level of Pathology	90	0	10
Description of the conduction that therapy was administered under.	50	30	20
Description of Placebo Used	10	70	**

\*\* Two other studies use placebo group, However the placebo group was combined with another treatment condition. It was not a "pure" placebo group.

Table 2  
Studies that met each criterion.

Study	Stat.	Internal Construct						External		Additional.					
		1	1	2	3	4	5	6	1	1	2	1	2	3	4
Criterion		1	1	2	3	4	5	6	1	1	2	1	2	3	4
Bassett & Pilowsk	y	x	y	x/y	y/y	y/y	y	x	y/y	y	0	y	x		
Beck, et al.	y	y	y	x/y	y/y	y/y	y	x	y/y	y	0	y	0	x	
Dedman, et al.	y	x	y	y/y	y/y	y/x	y	x	y/y	y	y	y	x		
Kirkley, et al.	y	x	y	x/y	y/y	y/x	y	x	y/y	y	y	x	x		
Michelson & Mavissakalian	y	y	y	y/y	y/x	y/y	y	y	y/y	y	y	y	x		
Mitcehl, et al.	x	x	y	x/y	y/y	y/y	y	y	y/y	y	y	x	y		
Moore & Chaney	x	y	y	x/y	y/y	y/y	y	x	y/y	y	y	y	x		
Ordman & Kirschenbaum	y	y	y	y/y	y/y	y/y	y	x	y/y	y	y	x	x		
Shea, et al.	y	x	y	y/y	y/y	x/x	y	y	y/y	y	y	0	x		
Woody, et al.	y	y	y	y/y	y/y	y/0	y	x	y/y	y	y	y	x		

Y= Yes

X= No

0= Not Clear

\*\*\* Some criterion Had more than one section, this is the reason for the reason for the double critique, (eg. Y/X). See Appendix 2 for critique.

Internal validity

## History Effect

1. Fifty percent of the studies used patients who had been previously treated for the same condition. As a result of their experience in past treatment, they may have performed differently.

## Testing Effect

2. Researchers in 50 percent of the studies tested the patients more than one time on the same test that was used to take the final measure.

## Instrumentation Effect

3. The researchers of 10 percent of the studies did not use the same testing procedure for all patients examined. Ordman and Kirchenbaum, (1985) examined cognitive-behavioral therapy for the treatment of bulimia. Two treatment conditions were used, Brief-intervention and Full-intervention therapy. Patients' in the Brief-intervention condition were tested by one researcher before therapy and after therapy. Patients' in the Full-intervention treatment condition were tested weekly by graduate students. It was obvious that the testing procedure was more lenient for the Brief-intervention treatment condition.

### Selection Bias

4. The researchers in 10 percent of the studies used patients who had different symptoms. In addition, twenty percent of the researchers used patients that were selected from different populations. For example, Beck, et al. (1985) examined cognitive therapy for the treatment of depression. Some of the patients' treated suffered from other disorders, besides depression. In addition, some patients' were referred by other treatment professionals, and others were self referred. See Appendix 2 for complete evaluation of the studies.

### Construct validity

Construct validity may have been affected because thirty percent of the studies did not use a placebo group and/or used only one treatment condition to account for effects measured (ie. mono-operational bias). Three studies used a placebo group. However, Mechelson and Mavessakalian, (1985) combined their placebo treatment with three other modes of therapy. Mitchell et al. (1990) used the placebo group as a control for their medication treatment condition. It was not used as a control for their psychotherapy treatment condition. Mitchell, (1990) was the only study that used a

placebo group and provided sufficient information so that other researchers could replicate the group, if they wish.

#### External validity

The results indicate that nine of the ten studies used a standard mode of therapy, as well as patients with clinical symptoms (See Table 3). However, 50 percent of the studies did not provide enough information on the conditions under which therapy was administered so that other researchers could replicate the study from the information provided.

For example, Shea, et al. (1990) reported that " The treatments included cognitive-behavioral therapy, interpersonal therapy, imipramine plus clinical management, and placebo plus clinical management. All treatments were 16 weeks in length, with 16-20 sessions."

p. 712. This was the only information provided on the treatment conditions used.



Table 3

Level of pathology of subjects and mode of therapy used by researchers.

<u>Study</u>	<u>Pathology</u>	<u>treatment</u>
Bassett & Pilowsky	Pain	Psychodynamic Cognitive therapy
Beck, et al.	Depression	Cognitive therapy
Dedman, et al.	Bulimia Nervosa	Cognitive- Behavior
Kirkley, et al.	Bulimia Nervosa	Cognitive- Behavior
Michelson & Mavissakalian	Agoraphobia	Behavioral therapy
Mitchell, et a.	Bulimia Nervosa	Behavior & Cognitive- behavior
Moore & Chaney	Chronic Pain	Cognitive- Behavior
Ordman & Kirschenbaum	Bulimia Nervosa	Cognitive- Behavior
Shea, et al.	Depression	Cognitive- behavior

		Interpersonal therapy
Woody, et al.	Sociopathy	Supportive- Expressive & Cognitive- Behavior

\*\* Only modes of psychotherapy are listed per study. Some studies used other forms of therapy, (ie. drug therapy). They were not listed because only psychotherapeutic treatment was examined.

## Chapter IV

## Discussion

The first part of this chapter presents implications of the review. The second section describes the limitations of the review. The third section discusses suggestions for further research. The discussion focuses on the strengths and weaknesses of the present research on the effectiveness of psychotherapy. Implications for each type of validity, when they are affected, are discussed.

## Implications of the review

The results of this thesis suggest a need for improvement in the quality of research of psychotherapeutic outcome studies. The findings indicate that fundamental areas of validity are not adequately controlled for and/or were not taken into consideration by researchers.

Statistical Conclusion Validity

The findings indicate that 80 percent of studies examined used standardized tests to measure the effectiveness of therapy.

This is a strength of the present research on the effectiveness of psychotherapy. The validity and reliability of a standardized test has usually been

verified. Thus, when researchers use a standardized test to measure the effectiveness of psychotherapy, they can be relatively certain that the test will produce consistent measures with repeated application under the same conditions. This increases the probability that the test will accurately measure the magnitude of effect produced by therapy. Twenty percent of the studies examined did not use a standardized test to measure the effectiveness of therapy. When researchers use a test whose reliability has not been verified, the researcher cannot be sure that the data produced by the test is accurate. For example, the researcher cannot be sure that the same test, administered repeatedly under the same conditions, will produce the similar results. Thus, in situations where no effects are measured by the test, there would be uncertainty as to whether therapy had an effect and the test simply failed to record it. On the other hand, in situations where effects are measured, researchers could not be sure that the test accurately measured the magnitude of effect produced by therapy. Either situation may cause inaccurate accounts of the effectiveness of psychotherapy.

#### Internal Validity

The findings indicate, that of the studies examined, all researchers adequately controlled for maturation effect, all

researchers used tests that were cross-validated for the type of subjects examined, at least 90 percent of researchers adequately controlled for instrumentation effects, 90 percent of researchers used subjects with the same symptoms in all treatment groups, and all researchers randomly assigned subjects to treatment conditions. When researchers adequately control for maturation effect, they decrease the probability that changes in subjects (ie. such as growing older, wiser, taller, etc.) will affect the outcome to the study. When researchers use tests that are cross-validated for the type of subjects examined, changes measured by the test are likely to be accurate because the test is appropriate for the subjects being tested. When researchers adequately control for an instrumentation effect, results are less likely to be affected by changes in the instrument used to measure variations in subjects conditions. When subjects in all treatment conditions have the same symptoms, researchers can be relatively certain that characteristics of subjects, (ie. level of intelligence, income, maturity) do not cause one group to perform better than another group. When subjects are randomly assigned to treatment conditions, the probability of all treatment conditions being equal is increased. This allows for valid comparisons between groups. These are all

areas of strength in internal validity of the present research on the effectiveness of psychotherapy. When these areas of internal validity are adequately controlled for, the validity of cause and effect relationships, established between the implementation of treatment and subject's conditions, is increased.

The findings also indicate that 50 percent of researchers did not adequately control history effect, 50 percent of researchers repeatedly tested subjects with the same test used to measure the effectiveness of therapy, and 60 percent of researchers selected subjects from different populations. When researchers do not adequately control for history effect, events that occur between the beginning of treatment and the final measure of the effectiveness of therapy may affect the results of the study. When researchers test subjects more than twice on the test used to measure the effectiveness of psychotherapy, patients may alter their responses because they were familiar with the test.

Selection bias also occurs when researchers select subjects from different populations. For example, suppose a researcher examining Bulimia, selects half his patients from an eating disorder clinic, and the other half through newspaper and radio advertising. The researcher is likely to get a population of patients who believe they need

therapy and expect that therapy will help them. These patients' are likely to be more cooperative and respond better to therapy than patients who are skeptical of the benefits of therapy. Researchers are more likely to find positive results for therapy when they use patients who believe that therapy will benefit them.

These are all areas of weakness in internal validity of the present research on the effectiveness of psychotherapy. When these areas of internal validity are not adequately controlled for, researchers cannot be certain that changes in patients' conditions are entirely due to treatment intervention. Therefore, they cannot provide valid evidence for the research of support arguments, for or against the effectiveness of a particular form of psychotherapy.

#### Construct Validity

The findings indicate that, of the studies examined, 30 percent of researchers used more than one treatment condition, (eg. psychotherapy and control/placebo therapy) to account for variations in patients' condition. The researchers of these studies can be relatively certain that operations, specific to each treatment condition, caused variations in patients' conditions. On the other hand, 50 percent of researchers did not use more than one treatment condition to account for variations in patients' conditions.

It has been well documented that variables such as therapist experience, client-therapist compatibility, patient's attitude toward therapy, etc., all contribute to the effectiveness of therapy. Therefore, when researchers use only one treatment condition to account for effects measured, they are more likely to erroneously attribute improvement in the patients, when improvements are in actuality due to variables other than therapy, to therapy. Therapist, using only one treatment condition (ie. therapy) to account for the effects measured, may attribute improvements to therapy because they have used only therapy to account for measured effects. This affects the construct validity of studies because researchers cannot be certain that operations specific to therapy, and therapy alone, caused changes in patients' conditions. Thus, while treatment leads to improvement in patients' conditions, researchers cannot be sure what construct is the treatment.

#### External Validity

The findings indicate that, of the studies examined, all researchers adequately controlled for interaction of selection and treatment, and interaction of setting and treatment. These are strengths of external validity of the present research on the effectiveness of psychotherapy. Researchers of these studies expressed caution when



generalizing their findings across populations. These researchers did not generalize their findings to populations different from the population used in their study, (interaction of selection and treatment). Researchers of these studies also expressed caution and did not generalize their findings to settings different from the setting of which their study conducted, (interaction of setting and treatment).

#### The Mode of Therapy Used

The results suggest that the researchers of 90 percent of the studies examined, used a standard mode of therapy. This is a strength of the present research on the effectiveness of psychotherapy. When researchers use standard modes of therapy (psychoanalysis, behavior modification, cognitive therapy, etc.), it is much easier for other researchers to replicate the study to verify the findings because other researchers are likely to be skilled in administering standard modes therapy. Thus, they can replicate the treatment with relative precision. On the other hand, when researchers use therapy, other than a standard mode of therapy, to test the effectiveness of psychotherapy, it is difficult to replicate the study with the same precision because other researchers are not likely to be familiar with, or as skilled, at administering the therapy.

Level of Pathology

The findings indicate that researchers of 90 percent of studies that examined the effectiveness of psychotherapy use patients with clinical symptoms. This is a strength of the present research on the effectiveness of psychotherapy. When researchers use patients with clinical symptoms, they can generalize their findings to specific populations of patients. For example, when researchers use patients who meet an established criteria (ie. DSM-III) for the condition being examined, the findings can be generalized to all patients who meet the criteria for the same condition. On the other hand, when researchers use patients with subclinical concerns, (eg. patients' with mild symptoms of a given disorder) the findings may not generalize to other patients suffering from the same conditions. This is because these patients' may have more severe clinical symptoms than the patients used to test the effectiveness of the therapy.

Description of the Condition under which Therapy was Used

The findings indicate that researchers of 50 percent of the studies examined provided enough information on the conditions under which therapy was administered so other researchers could replicate the treatment from the information provided. On the other hand, 50 percent of

researchers did not provide enough information on the condition under which therapy was used so that other researchers could replicate the treatment. This is a weakness of the present research on the effectiveness of psychotherapy. The researchers of these studies merely reported the mode of therapy that was used, and in cases where the researchers described the therapy used, their description was vague and incomplete. For example, Mitchell (1990) treated bulimia patients in a three phase treatment approach. "Phase 1, the preparatory phase, subjects were seen for two 2-hour group sessions each week for 2 weeks." "Phase 2, the interruption phase, there was the explicit exception that group participants should attempt to interrupt their bulimia behaviors and to begin to eat regular balanced meals." "The last phase, or stabilization phase, included the last month of the short-term treatment program and involved a single 1 1/2-hour session each week." For each phase, the authors provide a vague description of what they focused on. (p. 150). In another study, Beck, et al. (1985) stated "The cognitive therapy component of treatment for both groups was based on a manual by Beck et al. that was later published as a monograph." p. 144. The manual of cognitive therapy by Beck et al. may adequately describe therapy, however, it is not clear how closely the

manual was followed by the researchers since no information on their treatment procedure was reported. It is important that researchers provide sufficient information on the mode of therapy used, and the condition under which therapy was administered, so that other researchers can, from the information provided, replicate the therapy, if they wish. When researchers do not provide enough information on the therapy used, it is difficult to replicate the study.

#### Description of Placebo Used

The findings indicate that researchers of 50 percent of studies that examine the effectiveness of psychotherapy did not use a placebo group. The results also suggest that only 10 percent of those studies provided sufficient information on the placebo group so that other researchers could replicate the study. This is a weakness of the present research on the effectiveness of psychotherapy. It is important that researchers use a placebo group when testing the effectiveness of psychotherapy. According to Senger, (1987) "Placebo has three main components: inputs from the patient, therapist, and treatment." p. 76. The expectation of patients (ie. input from the patients) affects how well therapy works. The expectation of therapists (ie. input from the therapist) affects how well therapy works, and the compatibility of treatment, patient and therapist, (ie.

input from the treatment) affects how well therapy works. Researchers must control for each of these components because they influence the performance of psychotherapy. Senger argues that "Repeating quantitative measures of patient perception of these variables in reference to expectation and credibility of the treatment and relationship in all components represent a minimum requirement in any attempt to show incremental effectiveness of a psychotherapy." (p. 67). According to Critelli & Neumann, (1984) "Too often in the past, false claims of incremental effectiveness of therapy have resulted from the experimental use of placebos that even the most naive would not mistake for genuine therapy. There appears to be a tendency for experimental placebos to be in some sense weaker, less credible, or applied in a less enthusiastic manner than treatments that have been offered as actual therapies." (p.38). When researchers do not use a placebo, they cannot rule out the probability that placebo effect was responsible for effects measured. Thus, researchers cannot be certain of what is producing the treatment effect.

In summary, the results of this thesis indicate a possibility that all four types of validity, of the research on the effectiveness of psychotherapy, are affected by some of the threats to validity. The evidence from this thesis

## Validity

suggests a need for improvement in the research practice of efficacy studies, specifically the need for better control for some of the threats to the four types of validity. Furthermore, the evidence of this thesis suggests a need for better data reporting, specifically, that researchers need to report more information on their research procedures so that the study can be replicated to verify the findings. This author recognizes that it may not be possible to control for all threats to the four types of validity, (i.e. History, Maturation, Interaction of History and Treatment), however, in situations where there is a high probability that the results of a study is affected by a particular threat, researchers should report that probability and state why.

## Limitations of the review

A limitation of this review was the size of the sample. Since only ten studies were selected, there was a possibility that the sample was not representative of all types of studies that examine the effectiveness of psychotherapy.

## Suggestions for further research

Further research is needed to examine other threats to validity which may affect the findings of studies that examine the effectiveness of psychotherapy. As mentioned earlier, studies in psychology rarely examine the validity of research that report data for or against the effectiveness of psychotherapy. More research is needed to thoroughly examine the validity of the research on the effectiveness of psychotherapy. The following studies are suggested for further research:

1. An examination, similar to this review, using a larger sample.
2. Conducting pre-evaluations of validity of studies used in meta-analyses. Since one of the criticisms of meta-analyses is that they are based on studies of diverse quality, pre-evaluation. To make sure that all studies meet certain levels of validity, may eliminate this criticism, and strengthen the argument of meta-analyses.

## Reference

- Bassett, D. L., & Pilowsky, I. A study of Brief Psychotherapy For Chronic Pain, Journal of Psychosomatic Research, 29, 259-264.
- Beck, A. T., Hollon, S. D., Young, J. E. Bedrosian, R. C., & Budenz, D. (1985) Treatment of Depression With Cognitive Therapy and Amitriptyline. Archives of General Psychiatry 42, 142-148.
- Cook, T. D., & Campbell, D. T. (1977) Quasi-experimentation Design: Design and Analysis for Field Setting. Chicago, Il. Ran McNally.
- Dedman, P. A., Numa, S. F., & Wakeling (1988) A Cognitive Behavioral Group Approach For The Treatment of Bulimia nervosa- A preliminary study. Journal of Psychosomatic Research 32, pp. 285-290.
- Eysenck, H. J. (1952). The effects of Psychotherapy: an evaluation. Journal of Consulting Psychology, 16, 319-324.
- Kirk, G. E. (1982). Experimental design: Procedures for Behavioral Science (2nd ed.). Belmont, CA. Brook/co.
- Kirkley, B. G., Schneider, J. A., Agras, W. S., & Bachman, J. A. (1985) Comparison of Two Group Treatments for Bulimia. Journal of Consulting and Clinical Psychology. 53, 43-48.



- Michelson, L., & Mavissakalian, M. (1985)  
Psychophysiological Outcome of Behavioral and  
Pharmacological Treatments of Agoraphobia. Journal of  
Consulting and Clinical Psychology, 53, 229-236.
- Mitchell, J. E., Pyles, R. L., Eckert, E. D., Hatuskami, D.,  
Pomeroy, C., & Zimmermann, R. (1990) A comparison Study  
of Antidepressants and Structured Intensive Group  
Psychotherapy in the Treatment of Bulimia Nervosa.  
Archives of General Psychiatry, 47, 149-157.
- Moore, J. E., & Chaney, E. F. (1985) Outpatient Group  
Treatment of Chronic Pain: Effects of Spouse Involvement.  
Journal of Consulting and Clinical Psychology, 53, 326-  
334.
- Ordman, A. M., & Kirschenbaum, D. S. (1985) Cognitive-  
Behavioral Therapy for Bulimia: An Initial Outcome Study.  
Journal of Clinical Psychology, 53, 305- 313.
- Shapiro, D. A. & Shapiro, D. (1983) Comparative Therapy  
Outcome Research: Methodological Implications of Meta-  
Analysis. Journal of Consulting and Clinical Psychology,  
51, 42-53.
- Singer, H. R. (1987). The "Placebo" Effect of  
Psychotherapy: A Moose in the Rabbit Stew. American  
Journal of Psychotherapy, 31, 68-81.

- Shea, M. T., Pilkonis, P. A., Beckham, E., Collins, J. F., Elkin, I., Sotsky, S. M., & Docherty, J. P. (1990) Personality Disorders and Treatment Outcome in the NIMH Treatment of Depression Collaborative Research Program. American Journal of Psychiatry, 147, 1990.
- Smith M. L., & Glass, G. (1977). Meta-analysis of Psychotherapy Outcome Studies. American Psychologists, 32, 752-760.
- Smith, M. L., Glass, G. & Miller, T. I. (1980). The benefits of Psychotherapy. Baltimore: Johns Hopkins University Press.
- Woody, G. E., McLellan, A. T., Luborsky, L., & O'Brian, C. P. (1985) Sociopathy and Psychotherapy Outcome. Archives of General Psychiatry, 42, 1985.

## Appendix 1

Bassett, D. L., & Pilowsky, I. A study of Brief

Psychotherapy For Chronic Pain, Journal of Psychosomatic Research, 29, 259-264.

Twenty-six patients, suffering from chronic pain, were treated with either 12 sessions of psychodynamic therapy or six sessions of cognitive supportive therapy. The patients included seventeen females and five males, all between 22-55 years of age. The patients were randomly assigned to either one half hour sessions of supportive therapy every fortnight or twelve, weekly, one hour, sessions of dynamic therapy. Measures were taken at the completion of treatment, and at 6 and 12 month follow-up periods. Bassett and Pilowsky found that patients who received psychodynamic therapy reported significantly greater improvement than patients who received supportive therapy.

Beck, A. T., Hollon, S. D., Young, J. E. Bedrosian, R. C., & Budenz, D. (1985) Treatment of Depression With Cognitive Therapy and Amitriptyline. Archives of General Psychiatry 42, 142-148.

Beck et al. examined the effectiveness of cognitive therapy alone, and cognitive therapy plus amitriptyline hydrochloride pharmacotherapy for the treatment of primary nonbipolar depression. Nine men and twenty-four women were

randomly assigned to the two treatment conditions for 20 sessions of therapy over a 12 week period. Beck et al. found that, for the treatment of patients with primary nonbipolar depression with cognitive therapy alone, the addition of a tricyclic antidepressant does not increase the patient's response to cognitive psychotherapy alone.

Dedman, P. A., Numa, S. F., & Wakeling (1988) A Cognitive Behavioral Group Approach For The Treatment of Bulimia nervosa- A preliminary study. Journal of Psychosomatic Research 32, pp. 285-290.

Eight patients who met the DSM-III criteria for bulimia were selected in chronological order from a hospital's waiting list. The patients, all female between 18-26 years of age, reported bingeing an average of 14 times per week. The patients received weekly sessions of group cognitive behavior therapy for 15 weeks. Dedman, et al. found a decrease in bingeing episodes per week, and a decrease in the level of depression and anxiety experienced by the patients.

Kirkley, B. G., Schneider, J. A., Agras, W. S., & Bachman, J. A. (1985) Comparison of Two Group Treatments for Bulimia. Journal of Consulting and Clinical Psychology. 53, 43-48.

Twenty-eight women who met the DSM-III criteria for bulimia were treated with either cognitive-behavior therapy

or nondirective group therapy. The patients, all female between 18 and 46 years of age, reported bingeing and vomiting at least two times per week. The patients received weekly sessions of therapy over a 16 week period. Measures were taken at the completion of treatment, and 3 months after treatment. Kirkley, et al. found that cognitive-behavior therapy was more effective than nondirective therapy for the treatment of bulimia.

Michelson, L., & Mavissakalian, M. (1985)

Psychophysiological

Outcome of Behavioral and Pharmacological Treatments of Agoraphobia. Journal of Consulting and Clinical Psychology, 53, 229-236.

Sixty-two patients who met the DSM-III criteria for Agoraphobia were randomly assigned to one of four treatment conditions in a 2 X 2 factorial design. The patients were treated weekly with a combination of behavior therapy and pharmacotherapy over a 12 week period. Measures were taken on several different areas related to agoraphobia, from the severity of symptoms to heart rate. Measures were taken before treatment, at 4 weeks, 8 weeks, 12 weeks during treatment, and 1-month after the completion of treatment. Michelson and Mavissakalian found significant evidence for the effectiveness for the combination of behavior therapy

and pharmacotherapy for the treatment of agoraphobia.

Mitchell, J. E., Pyles, R. L., Eckert, E. D., Hatuskami, D., Pomeroy, C., & Zimmermann, R. (1990) A comparison Study of Antidepressants and Structured Intensive Group Psychotherapy in the Treatment of Bulimia Nervosa. Archives of General Psychiatry, 47, 149-157.

One hundred fifty-five outpatients who met the DSM-III criteria for bulimia were treated with either cognitive behavior therapy or drug therapy (tricyclic antidepressant). The patients, all female between 18 and 40 years of age, were randomly assigned to one of four treatment cells: imipramine hydrochloride treatment, placebo, imipramine plus intensive group therapy, and placebo plus intensive group therapy. The patients received treatment over a twelve week period. Mitchell, et al. found evidence for both antidepressant and structured psychotherapy for short-term treatment of bulimia.

Moore, J. E., & Chaney, E. F. (1985) Outpatient Group Treatment of Chronic Pain: Effects of Spouse Involvement. Journal of Consulting and Clinical Psychology, 53, 326-334.

Forty-three patients experiencing chronic pain for at least 6 months were treated with cognitive-behavior therapy. All patients, except one, were male between 23 and 69 years

of age. The patients were randomly assigned to one of three treatment conditions those being, couple group therapy, patient-only-group therapy, and waiting-list control. Moore and Chaney found patients in both treatment groups showed significant improvement over the patients in the waiting-list control group. There was no evidence that the patients' spouse involvement increased the patients' participation in therapy.

Ordman, A. M., & Kirschenbaum, D. S. (1985) Cognitive-Behavioral Therapy for Bulimia: An Initial Outcome Study. Journal of Clinical Psychology, 53, 305- 313.

Twenty bulimia patients who met the DSM-III criteria for bulimia were treated with cognitive-behavior therapy. The patients, all female between 18 and 30 years of age, were randomly assigned to either a Brief-intervention-waiting-list condition or a Full-intervention treatment condition. The patients were assessed on several areas related to bulimia. The areas ranged from eating attitudes to attitudes toward women. Ordman and Kirschenbaum found that patients who received Full-intervention-cognitive-behavior therapy had a significantly greater improvement than patients who received brief-intervention-waiting-list treatment condition.

Shea, M. T., Pilkonis, P. A., Beckham, E., Collins, J. F.,

Elkin, I., Sotsky, S. M., & Docherty, J. P. (1990)  
Personality Disorders and Treatment Outcome in the NIMH  
Treatment of Depression Collaborative Research Program.  
American Journal of Psychiatry, 147, 1990.

Two hundred and thirty-nine outpatients who met the Research Diagnostic Criteria (RDC) for depression were randomly assigned to one of four treatment groups. These groups were cognitive-behavior therapy, clinical therapy, imipramine plus interpersonal therapy imipramine plus clinical management, and placebo plus clinical management. Measures were taken at pretreatment, 4 weeks, 8 weeks, 12 weeks during treatment, and at treatment termination. Measures were taken on the frequency of personality disorders, attrition rates, depressive symptoms, and social and work functioning. Shea, et al. found that patients with personality disorders had more "notable personality disorders," (p. 713). Also, there was no evidence of depression for patients with personality disorders. Both showed improvement in social and work functioning.

Woody, G. E., McLellan, A. T., Luborsky, L., & O'Brian, C. P. (1985) Sociopathy and Psychotherapy Outcome. Archives of General Psychiatry, 42, 1985.

One hundred and ten nonpsychotic opiate addicts who met the DSM-III criteria for Antisocial Personality Disorder



were randomly assigned to one of three treatment conditions. These being drug counseling alone, supportive-expressive therapy plus counseling, or cognitive behavior therapy plus counseling. All patients were men between 18 and 55 years of age. Results were reported on four groups, Those being opiate dependence only, opiate dependence plus depression, opiate dependence plus depression plus antisocial personality disorders, and opiate dependence plus antisocial personality disorders. Woody, et al. found significant improvement in patients in all groups except opiate dependence plus antisocial personality disorders on most areas of assessment.

Appendix 2

Bassett, D. L., and Pilowsky, I. (1985).

CRITERIA: \_\_\_\_\_ YES NO

VALIDITY:

History Not Clear

Were the subjects previously treated for the same symptoms, and did the previous treatment caused them to have a positive or a negative view of therapy?

The researchers did not report information of the patient's treatment history. However, it would be logical a assumption that since the patients were referred by a pain clinic the patients were previously treated.

Maturation Not Clear

Did the researchers allow more time to elapse before taking the final measure that they allowed between treatment applications? (e.g. if the subjects received treatment one time a week, did the researchers allow more than that one week to elapse before taking the measure of the effectiveness of therapy?).

There does not appear to be any evidence of maturation. The researchers also reported

that the subjects were asked to complete, again the questionnaires administered prior to treatment at treatment termination." (p. 261).

Testing

No

1. Were the subjects tested repeatedly for the same effect or for different effects.

The patients were tested before and after treatment. "All assessments were reported six and twelve months after completion of treatment or withdrawal from therapy." (p. 261).

2. Is the test validated (and Cross-validated for the types of subjects used in the study?

Yes

The researchers used the Levine-Pilowsky depression Questionnaire LPD, and the Spielberger State-Trait Anxiety Inventory STAI, which are valid test, to assess different aspects related to pain. The Illness Behavior Questionnaire (IBQ) was the primary measure of pain related

symptoms. The (IBQ) provided measures on General Hypochondriasis, Disease Conviction, etc. The Illness Behavior Questionnaire is a valid test for pain symptoms. (p. 260).

Instrumentation

Yes

1. Was the testing procedure of the study consistent across subjects? (e.g. was every subject tested the same way and with the same test?).

"Patients who were assigned to the supportive psychotherapy group were treated differently from the patients who were assigned to the dynamic psychotherapy group. "Patients assigned for supportive psychotherapy ...emphasis was placed upon active involvement by the therapists, with questions, advice, and specific directions concerning the patients approach to their pain." In contrast, "patients receiving dynamic psychotherapy were instructed to verbalize freely with relatively little involvement by the

therapist." (p. 261). Also, patients in the supportive psychotherapy received half the amount of treatment that patients in the dynamic psychotherapy group received. (p. 260). However, the testing was consistent across the subjects.

2. Was the criteria of what constituted a therapeutic effect consistent across subjects? No

The measure of the effectiveness of therapy appeared to be based on the subjects rating themselves on a Global Assessment scale from, "Much worse to Much improved." (p 260).

Selection Bias No

1. Did the subjects in both the experimental and control group have different symptoms? No  
All patients had the same symptoms.
2. Were the subjects selected from different population? No  
"All subjects were referred from the pain clinic of a large metropolitan general hospital." (p. 260).

Randomization

Yes?

Were the subjects randomly assigned to the experimental and control groups?

"Patients were randomly allocated to either six fortnights half-hour sessions for supportive psychotherapy or twelve weekly one-hour sessions or dynamic psychotherapy."  
(p. 260).

Reliability of the Measure

Yes

was the measure reliable? Did the instrument produce consistent results with repeated testing? (e.g. if several measures were taken before the treatment was administered, would each measure be the same?).

The researchers used several tests that were valid and reliable tests. However, the primary dependent measure was a subjective questionnaire. The fact that patients are experiencing pain, this may affect how they rated themselves on the questionnaire.

Interaction of History and Treatment

No

Was the study conducted at a particular time and/or within a particular time period,

where the characteristics of that time would negatively or positively affect treatment? (e.g. was the study conducted within a year after the end of the Vietnam War?).

Dose not apply.

Interaction of Selection and Treatment No

1. Were the results of the study incorrectly generalized to a different population than was used by the study?  
 "Our findings indicated that patients who received twelve, hour-long sessions of dynamic psychotherapy tended to report improvement on a number of parameters as compared with groups receiving six half-hour sessions of supportive psychotherapy." (p. 263).

2. Did the researchers use "real patients," patients who were not solicited for participation in study, and who were not "perfectly healthy" college students? Yes

Even though the descriptions of the

characteristics of the patients were vague, it was clear that the were "real patients."

Interaction of Setting and Treatment No

Was the treatment condition and the control condition conducted in different settings and were the results incorrectly generalized across settings?

It appeared that the study was conducted at the University of Adelaid, Department of Psychiatry.

The Mode of Therapy Used No

Did the researchers use a standard mode of therapy? (e.g. psychoanalysis, rational emotive therapy, behavior therapy, etc.). Also, were those studies that used behavior therapy conducted over an extended period of time, (at least 5 sessions of therapy), and/or did they include a follow-up study?

"Psychodynamic Psychotherapy and cognitive oriented supportive psychotherapy." (p 260).

"Patients assigned for supportive psychotherapy were...placed upon active involvement by the therapists, with



questions, advice and specific directions concerning the patient's approach to their pain. Cognitive strategies...patients were generally encouraged to talk about any problems which concerned them and they wished to discuss." (p. 261). If this is an accurate description of the treatment procedure, this could hardly be considered a standard mode of therapy.

Level of Pathology

Not Clear

Did the subjects have subclinical concerns? (e.g. anxiety, self-esteem, assertiveness, etc.) The study must have used patients with some level of psychological pathology, preferably patients with severe psychological pathology.

No description of the patient population was reported.

The Description of Therapy Used

Yes

Did the researchers specify a description of therapy or give its criteria of what was considered psychotherapy. If so, was the description specific enough that other researchers could replicate the therapy, if

they wish?

The researchers gave an adequate description of the therapy used.

The Description of Placebo Used No

Did the researchers use a placebo group, and if so, did the study specify a description of placebo or give its criteria of what was considered a placebo? Also, was the description specific enough?

No placebo or control group was used.

The reliability of the primary measure of pain, the Illness Behavior Questionnaire, is questionable because it was based on self-reports by patients who were in pain, and individuals in pain are hardly objective. Furthermore, the sample population was not representative of chronic pain sufferers. The mode of therapy used was not a standard form of therapy with specific criteria. The researchers also did not use a control or placebo group. Thus, the practical application of this study is limited, specifically the results may not be generalizable to the unrestricted population. The results are also questionable because of the high attrition rate. If treatment was helping the patients, why were they leaving treatment? Also,

six and twelve months follow up assessments indicate that patients got better in the absence of treatment. Research shows that patients got worse as time pasted after treatment termination.

Validity 63

Beck, et al. (1985).

CRITERIA: YES NO

VALIDITY:

History No

Were the subjects previously treated for the same symptoms, and did the previous treatment cause them to have a positive or a negative view of therapy?

"Ten patients receiving cognitive therapy alone and 9 receiving the combined therapy had been treated previously with tricyclics." p. 143. The researchers also reported that "The proportion of subjects who showed any indication of prior knowledge of cognitive therapy, whether by word of mouth or through familiarity with literature, was comparable for the two groups; five (27%) of the therapy alone group and four (26%) of the combined group showed some indication of potential expectation biases." (p. 143).

Maturation No

Did the researchers allow more time to elapse before taking the final measure that they allowed between treatment applications? (e.g. if

the subjects received treatment one time a week, did the researchers allow more than that one week to elapse before taking the measure of the effectiveness of therapy?).

"At treatment termination, the full intake assessment battery and a thorough clinical evaluation were readministered." (p. 144).

### Testing

No

1. Were the subjects tested repeatedly for the same effect or for different effects?

"Patients were evaluated within seven days of telephone screening by an experienced psychiatrist or psychologist." (p. 142). The subjects were evaluated again at treatment termination. However, because the tests were administered 12-weeks apart it is not likely that the subjects would have become so familiar with the test the first time that it would affect their scores 12-weeks later. (p. 142-143).

2. Is the test validated (and cross-

Yes

validated), for the type of subjects used in the study?

Beck Depression Inventory (BDI)

Hamilton Rating Scale for Depression (HRSD).

### Instrumentation

Yes

1. Was the testing procedure of the study consistent across subjects? (e.g. was every subject tested the same way and with the same test?)

All subjects were screened using the same screening procedure, and it appeared that the subjects were tested using the same testing procedure. (p. 142-143).

2. Was the criteria, of what constituted a therapeutic effect, consistent across subjects?

Yes

Even though no specific criteria was laid out by the researchers, it was clear that a decrease in the frequency of depressive symptoms was the criteria for the effectiveness of therapy.

Selection Bias

Yes

1. Did the subjects, in both the experimental and control group, have different symptoms?  
Even though all of the subjects met the DSM-II criteria for diagnosis of depressing neurosis, several patients had addition personality disorders. (p. 144).

2. Was the subjects selected for different populations?  
All subjects were either self-referred or professional referred. (p. 142).

Yes

Randomization

Yes

Were the subjects randomly assigned to the experimental and control groups?  
The researchers did not use a control group. However, "Patients were randomly assigned to treatment groups: 18 to cognitive therapy and 15 to combined cognitive therapy and pharmacotherapy." (p. 144).

Reliability of the Measure

Yes

Was the measure reliable? Did the instrument produce consistent results

with repeated testing? (e.g. if several measures were taken before the treatment was administered, would each measure be the same?).

Beck Depression Inventory

Hamilton Rating Scale for Depression.

Interaction of History and Treatment No

Was the study conducted at a particular time and/or within a particular time period, where the characteristics of that time would negatively or positively affect treatment? (e.g. was the study conducted within a year after the end of the Vietnam War?).

Does not apply.

Interaction of Selection and Treatment Yes

1. Were the results of the study incorrectly generalized to a different population than as used by the study?

"The results of the present researchers suggest that, treating outpatients with nonbipolar depression similar to those described herein, the addition of a tricyclic antidepressant does not



seem to add appreciably to the good short-term response associated with cognitive therapy. In light of the side effects and adverse reactions many patients experience with tricyclic antidepressants, it seems that cognitive therapy alone is an effective alternative to tricyclic." (p. 148).

2. Did the researchers use "real patients," patients who were not solicited for participation in the study, and who were not "perfectly healthy" college students? Yes

All subjects met the DSM-II criteria for diagnosis of depressive neurosis. p. 144. However, some of the subjects were "self-referred" to the study and it was not clear how they came to know about the study.

Interaction of Setting and Treatment

No

Was the treatment condition and the control condition conducted in different settings, and were the results incorrectly generalized across settings?

The study was conducted at the Mood Clinic of the Center for Cognitive Therapy, which is part of the Hospital of the University of Pennsylvania, Philadelphia, and the results were not generalized across settings.

The Mode of Therapy Used

Yes

Did the researchers use a standard mode of therapy? (e.g. psychoanalysis, rational emotive therapy, behavior therapy, etc.). Also, were those studies that used behavior therapy conducted over an extended period of time, (at least 5 sessions of therapy), and/or did they include a follow-up study?

The researchers used cognitive therapy and cognitive therapy combined with amitriptyline. (p. 144). Also, "Therapist had at least six months of training and supervised experience with cognitive therapy prior to treating their first study patients." (p. 144).

Yes, the study was conducted over a 12-week period with follow-up evaluations.

Level of Pathology

Yes

Did the subjects have subclinical concerns?

(e.g. anxiety, self-esteem, assertiveness, etc.) The study must have used patients with some level of psychological pathology, preferably patients with severe psychological pathology.

All subjects met the DSM-II criteria for depression and, according to the history of illness and characteristics of the patients reported by the researchers, it appeared that the subject's symptoms were at the clinical level. (See table 2. p. 143).

The Description of Therapy Used

Not clear

Did the researchers specify a description of therapy or give its criteria of what was considered psychotherapy? If so, was the description specific enough that other researchers could replicate the therapy, if they wish?

The researchers referred to a treatment manual. However the manual is just a treatment format that does not specify a criteria of treatment. "The cognitive therapy component of treatment for both was based on a manual by Beck et al that was

later published as a monograph. The researchers description of the treatment process of the group receiving combined cognitive therapy and pharmacotherapy, was more detailed." (p. 144).

The Description of Placebo Used No

Did the researchers use a placebo group, and if so, did the study specify a description of placebo or give it's criteria of what was considered a placebo? Also, was the description specific enough?

No, placebo or control group was used.

The researchers reported that some (26% to 27%) of the patients showed some indication of prior knowledge of the therapy administered which showed some potential expectation bias. The sample population did appear to be bias as a result of the "self-referral" to the subjects. Furthermore, the researchers did not use a placebo or a control group to control for extraneous factors. This study did not satisfy many of the criteria to present a valid argument for or against the effectiveness of treatment for depress patients.

Dedman, et al. (1988).

CRITERIA: YES NO

VALIDITY:

History Not Clear

Were the subjects previously treated for the same symptoms, and did the previous treatment cause them to have a positive or a negative view of therapy?

The study did not report the patient's previous treatment history. (p. 286).

Maturation No

Did the researchers allow more time to elapse before taking the final measure that they allowed between treatment applications? (e.g. if the subjects received treatment one time a week, did the researchers allow more than that one week to elapse before taking the measure of the effectiveness of therapy?)

In addition to being tested throughout treatment, the patients were tested at the end of treatment. 286.

Testing Yes

1. Were the subject tested repeatedly for the same effect or for different

effects?

Measures were taken before treatment, at week 7 of treatment, at the end of treatment period and at 3 and 6 months follow up." (p. 286).

2. Is the test validated (and cross-validate for the types of subjects used in the study? Yes

The Eating Attitudes Test (EAT), was used to measure attitudes towards food and dieting. This measure appeared to be a valid measure for this purpose. The researchers also used other tests to measure related factors of bulimia. The measures also appeared to be valid.

### Instrumentation

Yes

1. Was the testing procedure of the study consistent across subjects? (e.g. was every subjects tested the same way and with the same test?).

Both the treatment process and the different phases of treatment were consistent across patients.

2. Was the criteria of what constituted a therapeutic effect consistent across subjects?

No criteria was specified, however it was clear that awareness of the function of bulimic behavior and achieving a "normal," (three meals per day) eating habit was the measure of the effectiveness of therapy.

Selection Bias No

1. Did the subjects in both the experimental and the control group have different symptoms?

It appeared that all of the subjects had the same symptoms.

2. Were the subjects selected from different populations? No

"Patients were selected from GP referrals to the Academic Department of Psychological Medicine at the Royal Free Hospital." (p. 286).

Randomization No

Were the subjects randomly assigned to the experimental and control groups?

Validity 75

Patients were taken in chronological order from the waiting list. (p. 286).

Reliability of the Measure No

Was the measure reliable? Did the instruments produce consistent results with repeated testing? (e.g. if several measures were taken before the treatment was administered, would each measure be the same?).

The measure was based on "Self-reported frequency of binging and vomiting." (p.286). The self-report was based on a weekly estimate of each subject with reference to daily diary." The reader is reminded that self-reported measures are affected by the patients view as to his/her need for therapy, and the patients position on the benefits of therapy.

Interaction of History and Treatment No

Was the study conducted at a particular time and/or within a particular time period, where the characteristics of that time would negatively or positively affect treatment? (e.g. was the study conducted



within a year after the end of the Vietnam War?).

Does not apply.

Interaction of Selection and treatment No

Were the results of the study incorrectly generalized to a different population than was used by the study?

The researchers conclude that "Our approach fulfill some of these criteria as it of limited duration and is suitable for use by trained members of a variety of health care professions." p. 289.

2. Did the researchers use "real patients," patients who were not solicited for participation in the study and who were not "perfectly healthy" college students? Yes

"The mean duration of bulimic syndromes was 7.13 years (range 2-15 years) and mean weight was 96% (range 82-112% of the matched population mean weight, MPMW)." (p. 286).

Validity 77

Interaction of Setting and Treatment No

Was the treatment condition and the control condition conducted in the same setting, and were the results incorrectly generalized across settings?

The study was conducted at the Academic Department of Psychological Medicine at the Royal Free Hospital, and the results were not generalized across settings.

The Mode of therapy used Yes

Did the researchers use a standard mode of therapy? (e.g. psychoanalysis, rational emotive therapy, behavior therapy, etc.). Also, were those studies that used behavior therapy conducted over an extended period of time (at least 5 sessions of therapy), and/or did they include a follow-up study?

"Treatment employed a variety of technique derived from Behavior therapy and Cognitive therapies..." (p.286). Behavioral techniques were used to shape the patients to eat three meals per day. Cognitive techniques were used to help the patients recognize distorted thoughts related to eating habits.

Yes, "treatment was administered weekly for 15 sessions each of one and a half hour duration." (p. 286).

Level of Pathology

Yes

Did the subjects have subclinical concerns? (e.g. anxiety, self-esteem, assertiveness, etc.) The study must have used patients with some level of psychological pathology, preferably patients with severe psychological pathology.

Due to the mean duration of bulimic symptoms of 7.13 years, the patients appears to have have an adequate level of pathology for the purpose of the study. (p. 286).

The Description of Therapy Used

Yes

Did the researchers specify a description of therapy or give its criteria of what was considered psychotherapy. If so, was the description specific enough that other researchers could replicate the therapy, if they wish.

The therapist used a combination of behavior and cognitive techniques to treat the patients. The researchers did give a description of

the treatment used. "Treatment employed a variety of techniques derived from behavior therapy and cognitive therapies, and was divided into two phases." (p. 286).

The Description of Placebo Used

No

Did the researchers use a placebo group, and if so, did the study specify a description of placebo or give its criteria of what was considered a placebo? Also, was the description specific enough?

No placebo or control group was used. "We recognize that our study was uncontrolled, but felt that our results are promising enough to merit further pursuance of similar controlled treatment studies."  
(p. 286).

The study did not meet several of the criteria laid out in this thesis. Subjects were not randomly assigned to the treatment groups, they were repeatedly tested, and particularly with this population, repeated testing is a problem. The reliability of the measure is questionable since it was based on self-reporting. Self-reporting is also a problem with this population because these individuals have a distorted perception of their body image.

Also, the study did not use a placebo or a control group, therefore, the findings cannot be generalized to the unrestricted population of bulimic sufferers.

Validity 81

Kirkley, et al. (1985).

CRITERIA: \_\_\_\_\_ YES NO

VALIDITY:

History

Not Clear

Were the subjects previously treated for the same symptoms, and did the previous treatment caused them to have a positive or a negative view of therapy?

The researchers did not report information concerning the subjects previous treatment history.

Maturation

No

Did the researchers allow more time to elapse before taking the final measure that they allowed between treatment applications? (e.g. if the subjects received treatment one time a week, did the researchers allow more than that one week to elapse before taking the measure of the effectiveness of therapy?)

"Measures were taken one week prior to treatment and one week after treatment was terminated." (p. 44).

Testing

No

1. Were the subjects tested repeatedly for the same effect or for different effects?

All participants completed an eating history questionnaire prior to treatment to determine the duration and severity of bulimic behavior. They used standardized food records to monitor their eating and vomiting for 1 week prior to and 1 week following treatment." (p. 44).

2. Is the test validate (and cross-validated for the types of subjects used in the study?

Yes

The Eating History Questionnaire is a standardized test and is a valid test for evaluating bulimic disorder.

Instrumentation

Yes

1. Was the testing procedure of the study consistent across subjects? (e.g. was every subject tested the same way and with the same test?)

Since the study is comparing the two

groups, there is no treatment effect, what we have is a statistical effect. The procedure, of what constituted a statistical effect, was consistent across groups.

2. Was the criteria of what constituted a therapeutic effect consistent across subjects? Yes

Even though the researchers did not specify the criteria of what constituted a therapeutic effect, it appeared that the goal of therapy was to help the subjects stop bingeing and vomiting. (p. 45).

#### Selection Bias

No

1. Did the subjects in both the experimental and control group have different symptoms?

All subjects in both groups met the DSM-III criteria for the present condition.

2. Was the subjects selected for different populations? No

"All subjects were selected through news-



paper and television announcements  
describing the research project." p. 44.

Randomization

Yes

Were the subjects randomly assigned to the  
experimental and control groups?

"Those women whose monitoring indicated  
that they were vomiting between 2 and 50  
times per week were assigned to one of two  
treatment conditions (cognitive-behavior  
and nondirective) using the minimization of  
different technique to match the groups on  
vomiting frequency." (p. 45).

Reliability of the Measure

Yes?

Was the measure reliable? Did the  
instrument produce consistent results  
with repeated testing? (e.g. if several  
measures were taken before the treatment  
was administered would each measure be  
the same?).

The researchers used several tests that were  
reliable test such as the Beck and the  
Spielberger. However, the primary dependent  
measure was based on self-reporting in order  
to obtained data. For obvious reasons

the reliability of data, obtained through self-reporting, from clients who are afflicted by this disease are questionable.

Interaction of History and Treatment

No?

Was the study conducted at a particular time and/or within a particular time period, where the characteristics of that time would negatively or positively affect treatment? (e.g. was the study conducted within a year after the end of the Vietnam War?).

However, the study was conducted in 1985, and the 80's era was an era when body image was a primary concern and little was known about the disorder. People were not very sympathetic to bulimic sufferers. Therefore, the supportive climate that patients needed to help increase the chance that therapy would be successful, may not have been present. This would have affected the performance of the clients in treatment.

Interaction of Selection and Treatment

Yes

1. Were the results of the study incorrectly generalized to a different population

than as used by the study?

"Results of the present study indicate that group treatment for bulimia can be effective...a cognitive-behavior focusing on specific behavior changes yields results superior to less directive approaches." (p. 46).

2. Did the researchers use "real patients," patients who were not solicited for participation in the study, and who were not "perfectly healthy" college students? Yes

All subjects met the DSM-III criteria for bulimia. (p. 44).

Interaction of Setting and Treatment

No

Was the treatment condition and the control condition conducted in different setting, and were the results incorrectly generalized across settings?

It appeared that the study was conducted at Stanford University School of Medicine.

The Mode of Therapy Used

No

Did the researchers use a standard mode of therapy? (e.g. psychoanalysis, rational

emotive therapy, behavior therapy, etc.).

Also, were those studies that used behavior therapy conducted over an extended period of time (at least 5 sessions of therapy), and/or did they include a follow-up study?

The researchers used cognitive-behavior therapy and nondirective treatment. No

However, there are no clear criteria of what constitutes nondirective treatment.

Yes, the study was conducted over a 16 week period.

Level of Pathology Yes

Did the subjects have subclinical concerns? (e.g. anxiety, self-esteem, assertiveness, etc.) The study must have used patients with some level of psychological pathology, preferably patients with severe psychological pathology.

All subjects met the DSM-III criteria for Bulimia. (p. 45). "In addition, all the participants reported self-induced vomiting at least twice each week." (p. 44).

Validity 88

The Description of Therapy Used No

Did the researchers specify a description of therapy or give its criteria of what was considered psychotherapy. If so, was the description specific enough that other researchers could replicate the therapy, if they wish?

The researchers gave a description of therapy. However, the description they gave for the nondirective was vague, and there is no set criteria for nondirective therapy.

The Description of Placebo Used No

Did the researchers use a placebo group, and if so did the study specify a description of placebo or give its criteria of what was considered a placebo? Also, was the description specific enough?

No placebo group was used.

This study was not constructed very well. They did not satisfy several of these basic criteria. The reliability of the dependent measure is questionable because the measure was based on self-reporting by the patients. There was no set criteria for nondirective therapy. Also, the study did not use a placebo group.

Therefore, in addition to the results being questionable, they could not be generalized to the unrestricted population.

Validity 90

Mechelson, L. and Mavissakalian, M. (1985).

CRITERIA: YES NO

VALIDITY:

History

Yes

Were the subjects previously treated for the same symptoms, and did the previous treatment cause them to have a positive or a negative view of therapy?

The researchers report that 98% of the subjects had previously sought help for their agoraphobia, and 74% had received previous psychiatric treatment of an average duration of 25 months with little or no reported benefit. Because these subjects were previously treated, and treatment did not benefit them much, they are likely to have some preconceptions about treatment. This is particularly true of patients who are treated with pharamological treatment, and many of the subjects had received previous pharamological treatment. (p. 230).

Maturation

No

Did the researchers allow more time to elapse before taking the final measure that they

allowed between treatment applications? (e.g. if the subjects received treatment one time a week, did the researchers allow more than that one week to elapse before taking the measure of the effectiveness of therapy?)

Measures were taken at "4 weeks, 8 weeks, 12 weeks of treatment, and at 1-month posttreatment." (p. 230).

Testing

Yes

1. Were the subjects tested repeatedly for the same effect or for different effects?

Again, the subjects were tested at 4 weeks, 8 weeks, 12 weeks of treatment, and at 1-month posttreatment. (p. 230).

2. Is the test validated (and cross-validated) for the type of subjects used in the study?

Yes

The two primary measures of agoraphobia were a Standardized Behavioral Avoidance Course (S-BAC) and the Idiosyncratic Behavioral Avoidance Course (I-BAC). (p. 231).



Instrumentation

Yes

1. Was the testing procedure of the study consistent across subjects? (e.g. was every subject tested the same way and with the same test?)

All subjects were tested using the same assessment instruments and techniques.

All subjects received a behavioral assessment and a psychophysiological assessment. (p. 231).

2. Was the criteria of what constituted a therapeutic effect consistent across subjects?

Yes

Even though no specific criteria was laid out by the researchers, it was clear that a reduction of the level of anxiety, measured by heart rate, was the measure of the effectiveness of therapy.

Selection Bias

No

1. Did the subjects in both the experimental and control group have different symptoms?

All subjects met the DSM-III criteria for agoraphobia. In addition, the average

Validity 93

duration of agoraphobia was 10 years.

2. Was the subjects selected for different populations? No

It appeared that the study was conducted at the Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine.

Randomization Yes

Were the subjects randomly assigned to the experimental and control groups?

"The study employed a 2 (mediation/placebo) X 2 (flooding/discussion) factorial design with subjects randomly assigned to one of four conditions." (p. 230).

Reliability of the Measure Yes

Was the measure reliable? Did the instrument produce consistent results with repeated testing? (e.g. if several measures were taken before the treatment was administered, would each measure be the same?).

The subjects were administered a behavioral assessment and a psychophysiological assessment. For the behavioral assessment

the Standardized Behavioral Avoidance Course (S-BAC) and Idiosyncratic Behavioral Avoidance Course (I-BAC) were the primary measure of agoraphobia. These tests are reliable tests for measuring avoidance which is related to agoraphobia. The psychophysiological assessments were based on the patient's their heart rate. This is not a reliable measure of agoraphobia because there are several variables which affects ones heart rate. (p. 231).

Interaction of History and Treatment

No

Was the study conducted at a particular time and/or within a particular time period, where the characteristics of that time would negatively or positively affect treatment? (e.g. was the study conducted within a year after the end of the Vietnam War?).

This Criterion is not relevant to this study.

Interaction of Selection and Treatment

No

1. Were the results of the study incorrectly generalized to a different population than as used by the study?

The researchers summarized and discussed the results in terms of the different areas of the study. They did not generalize the results to the unrestricted population. (p. 231-234).

2. Did the researchers use "real patients," patients who were not solicited for participation in the study, and who were not "perfectly healthy" college students? Yes

All subjects met the DSM-III criteria for agoraphobia. In addition, the average duration of agoraphobia was 10 years. (p. 229-230).

Interaction of Setting and Treatment No

Was the treatment condition and the control condition conducted in different settings, and were the results incorrectly generalized across settings?

It appeared that the study was conducted at the Western Psychiatric Institute and Clinic University of Pittsburgh School of Medicine, and the results were not generalized across settings. (p. 229).

The Mode of Therapy Used

Yes

Did the researchers use a standard mode of therapy? (e.g. psychoanalysis, rational emotive therapy, behavior therapy, etc.). Also, were those studies that used behavior therapy conducted over an extended period of time (at least 5 sessions of therapy), and/or did they include a follow-up study?

The researchers used behavior therapy.

"Subjects were also given a comprehensive behavior rational for their conditions, emphasizing the role to habitual avoidance in maintaining their fears..." (p. 230).

Yes, the study was conducted over a 12-week period and was followed up by a 1-month posttreatment assessment. (p. 230).

Level of Pathology

Yes

Did the subjects have subclinical concerns? (e.g. anxiety, self-esteem, assertiveness, etc.) The study must have used patients with some level of psychological pathology, preferably patients with severe psychological pathology.

All subjects met the DSM-III criteria for

agoraphobia, and had an average duration of of agoraphobia of 10 years. (p. 229-230).

The Description of Therapy Used Yes

Did the researchers specify a description of therapy or give its criteria of what was considered psychotherapy? If so, was the description specific enough that other researchers could replicate the therapy, it they wish.

Even though no specific description of therapy was given by the researchers, the researchers did provide a clear description to the therapeutic instruction given to the subjects in each treatment group. (p. 230).

The Description of Placebo Used No

Did the researchers use a placebo group, and if so, did the study specify a description of placebo or give its criteria of what was considered a placebo? Also, was the description specific enough?

Even though the researchers did use a "placebo group" because it was combined with other treatments it is not a control.

"The control group was not an untreated

control condition but rather was specifically designed to address issues of therapeutic expectancy and not specific therapists/treatment factors and to equate contact time across all conditions."

(p. 230). If the placebo group was used as a separate treatment condition, then it could be used as a control for comparisons with other treatment conditions. However, because it was used in combination with other treatment conditions, it cannot be relied upon to control for extraneous variables.

(p. 230).

It was possible that this study was contaminated by history effect as a result of the subject's previous treatment history. It is also possible that this study was contaminated by testing effects as a result of the subjects being repeatedly tested through treatment. Furthermore, the study was not a true control evaluation. Therefore, the results cannot be generalized to the unrestricted population of agoraphobia sufferers.

Mitchell, J. E. (1990).

CRITERIA: YES NO

VALIDITY:

History

Not Clear

Were the subjects previously treated for the same symptoms, and did the previous treatment cause them to have a positive or a negative view of therapy?

The researchers did not report information concerning the subjects treatment history.

Maturation

No

Did the researchers allow more time to elapse before taking the final measure that they allowed between treatment applications? (e.g. if the subjects received treatment one time a week, did the researchers allow more than that one week to elapse before taking the measure of the effectiveness of therapy?).

All subjects that completed the study were evaluated at termination of the study, and subjects who did not complete the study, were evaluated at the time of their termination. The researchers performed analysis on all subjects who completed



Validity 100

5, 8, and 10 visits. However, "only the end point analysis (10 visit) were reported." (p. 150).

Testing

No

1. Were the subjects tested repeatedly for the same effect or for different effects?

All subjects were administered tests to obtain a baseline, and again at treatment termination. They were also evaluated at particular visits throughout treatment. (p. 150). However, particularly with this population repeated testing is a problem because these individuals have a distorted perception of themselves, and try to present what they believe is a "good image." With repeated testing, the subjects may become familiar with the test, and they may respond to the questions in a way that they feel may make them look good.

2. Is the test validate (and cross-validated) for the type of subjects

Yes

used in the study?

Eating Disorder Questionnaire

The Eating Disorder Inventory

instrumentation

Yes

1. Was the testing procedure of the study consistent across subjects? (e.g. was every subject tested the same way and with the same test?)

All subjects received the same test to obtain the baseline, and all were administered the same test at treatment termination. (p. 150).

2. Was the criteria of what constituted therapeutic effect consistent across subjects?

Yes

It was clear that a reduction in the number of binge-eating episodes per week, self-induced vomiting episodes per week, and time spent binge eating each week was the measure to the effectiveness of therapy.

Selection Bias

No

1. Did the subjects in both the experimental and control groups have

different symptoms?

"All subjects met the DSM-III criteria for bulimia, with the additional criterion of binge eating coupled with self-induced vomiting or laxative abuse at a minimum frequency of three times each week for the 6 months before evaluation." (p. 149).

2. Was the subjects selected from different populations? Yes

"Patients were recruited from the pool of patients being evaluated in the Eating Disorder Clinic at the University of Minnesota, Minneapolis, and from symptomatic Volunteers recruited....advertisements in local Newspapers and the radio." (p. 149).

Randomization

Yes

Were the subjects randomly assigned to the experimental and control groups?

"...Subjects who continued to satisfy admission criteria were stratified by the level of depression of the HDRS (score 15 vs 15) and randomized to one of

four treatment cells." (p. 150).

Reliability of the Measure

Yes?

Was the measure reliable? Did the instrument produce consistent results with repeated testing? (e.g. if several measures were taken before the treatment was administered, would each measure be the same?).

The researchers did use other tests that were reliable tests to assess other related aspects of the sample population. Even though the Eating Disorder Questionnaire and the Eating Disorder Inventory were the primary measures to eating disorder, and they were standardized tests. It should be noted that these test are based on self-reporting by the patients. Particularly with this population, self-reporting is a problem. The individuals cannot be relied on to accurately assess themselves because they have a distorted perception of their body image.

Interaction of History and Treatment No

Was the study conducted at a particular time and/or within a particular time period, where the characteristics of that time would negatively or positively affect treatment? (e.g. was the study conducted within a year after the end of the Vietnam War?).

The study was conducted at the Eating Disorder Clinic at the University of Minnesota, Minneapolis. The study was conducted in 1990. Unlike the 80's, much more is known about the disorder and there is more support for individuals who suffer from this disorder.

Interaction of Selection and Treatment Yes

1. Were the results of the study incorrectly generalized to a different population than as used by the study?

The researchers summarized the results in terms of the different comparisons. They also generalized the results to the general population. This generalization is appropriate because

the study did use a placebo group. However, because the sample population may not have a "true" representative sample, the validity of the results, when applied to the unrestricted population, is questionable. (p. 151-155).

2. Did the researchers use "real patients," patients who were not solicited for participation in the study and who were not "perfectly healthy" college students? Yes

All subjects met the DSM-III criteria for bulimia. However, "potential subjects were recruited from the pool of patients being evaluated in the Eating Disorders Clinic at the University of Minnesota, Minneapolis, and for symptomatic volunteers recruited for treatment studies at the University of Minnesota through advertisement in local newspapers and over the radio." (p. 149).

Interaction of Setting and Treatment No

Was the treatment condition and the control condition conducted in different settings, and were the results incorrectly generalized across settings.

The study was conducted in the Eating Disorder Clinic at the University of Minnesota, Minneapolis, and the results were not generalized across settings.

The Mode of Therapy Used Yes

Did the researchers use a standard mode of therapy? (e.g. psychoanalysis, rational emotive therapy, behavior therapy, etc.).

Also, were those studies that used behavior therapy conducted over an extended period of time (at least 5 sessions of therapy), and/or did they include a follow-up study?

The researchers used Behavioral and Cognitive behavioral techniques.

Level of Pathology Yes

Did the subjects have subclinical concerns? (e.g. anxiety, self-esteem, assertiveness, etc.) The study must have used patients with some level of psychological pathology,

preferably patients with severe psychological pathology.

All subjects met the DSM-III criteria for bulimia. (p. 149).

The Description of Therapy Used

No

Did the researchers specify a description of therapy or give its criteria of what was considered psychotherapy. If so, was the description specific enough that other researchers could replicate the therapy, if they wish?

The researchers merely reported what type of therapy was used. (p. 150).

The Description of Placebo Used

Yes?

Did the researchers use a placebo group, and if so did the study specify a description of placebo or give its criteria of what was considered a placebo? Also, was the description specific enough?

The placebo group was adequately described by the researchers. However, the way the placebo group was used only applied to the medication group in terms of making comparisons. The placebo group was not



constructed or used in such a way to permit a valid comparison with the psychotherapy group. The placebo administered was a pill. It seems to me, that if the placebo group is to serve as a comparison for the cognitive or any psychological treatment for that matter, it should be in a form similar to the psychological treatment procedure.

The study was well designed. However, due to the fact that some of the subjects were solicited or volunteered for the study and they came from different populations, the sample population may not be a "true" representative sample of bulimia sufferers. Also, due to the fact that the measures were based on self-reported data, the reliability of the findings are questionable. Furthermore, the researchers reported that the subjects lost weight but improved. Losing weight is not consistent with bulimia treatment.

Moore, J. E. and Chaney, E. F. (1985).

CRITERIA: YES NO

VALIDITY:

History

Yes

Were the subjects previously treated for the same symptoms, and did the previous treatment caused them to have a positive or a negative view of therapy?

"...9 patients had one surgery, 7 had had two surgeries, and 10 had three to nine surgeries." (p.327). Because the subjects had received previous treatments for the pain condition, It is likely that they may have preconceptions about the effects of therapy. This would affect how the performed in therapy.

Maturation

No

Did the researchers allow more time to elapse before taking the final measure that they allowed between treatment applications? (e.g. if the subjects received treatment one time a week, did the researchers allow more than that one week to elapse before taking the measure of the effectiveness of therapy?).

The measure of the effectiveness of therapy taken the week following completion of treatment. (p. 327).

Testing

No

1. Were the subjects tested repeatedly for the same effect or for different effects?

The subjects received a pretreatment assessment, the same assessment one the week following completion of treatment, and again at three months following completion of treatment. (p. 328).

2. Is the test validated (and cross-validated) for the types of subjects used in the study?

No

The researchers used other measures that were validated. However, the Visual Analogue Scale, was the primary measure of the subjects pain. (p. 327).

Instrumentation

Yes

1. Was the testing procedure of the study consistent across subjects? (e.g. was every subject tested the same way and

with the same test?)

In addition to being assessed, using the same procedure the subjects, the subjects were treated using the same treatment procedure with regard to their respective group. (p. 328).

2. Was the criteria of what constituted a therapeutic effect consistent across subjects? Yes

Even though no specific criteria was laid out by the researchers, it was clear that a reduction in the frequency of pain, experienced by the subjects, was the criteria for the effectiveness of treatment.

### Selection Bias

No

1. Did the subjects in both the experimental and control group have different symptoms?
- All patients appeared to have the same symptoms.
2. Was the subjects selected for different populations? Yes

"Patients were selected form any

referring hospital service (primarily orthopedic, neurosurgery, and rehabilitation medicine.)" (p. 327).

Randomization

Yes

Were the subjects randomly assigned to the experimental and control groups?

"Groups of 4-6 consecutively enrolled patients were randomly assigned to one of two treatment conditions (individual or couples treatment) or to a waiting-list control." (p.328).

Reliability of the Measure

No

Was the measure reliable? Did the instrument produce consistent results with repeated testing? (e.g. if several measures were taken before the treatment was administered would each measure be the same?).

The researchers did use several tests which were reliable tests to evaluate other aspects of the sample population. However, The Visual Analogue Scale (VAS) was the primary measure of pain. This test does appear to be a standardized or reliable

test for evaluation of pain. (p. 327-328).

Interaction of History and Treatment No

Was the study conducted at a particular time and/or within a particular time period, where the characteristics of that time would negatively or positively affect treatment? (e.g. was the study conducted within a year after the end of the Vietnam War?).

Does not apply.

Interaction of Selection and Treatment No

1. Were the results of the study incorrectly generalized to a different population than as used by the study?

The researchers summarized and discussed the results in terms of the different areas of the study and with in the context of the study. "The present study evaluated the efficacy of a brief outpatient group therapy program for chronic pain patients, and with the context of this program..." (p. 331).

2. Did the researchers use "real patients," patients who were not solicited for Yes

participation in the study, and who were not "perfectly healthy" college students?

"patients were referred by hospital referring service (primarily orthopedics, neurosurgery, and rehabilitation medicine) provided they had experienced pain for at least 6 months....," (p. 327).

Interaction of Setting and Treatment

No

Was the treatment condition and the control condition conducted in different settings, and were the results incorrectly generalized across settings?

The study was conducted at "Northwestern Veterans Administration general medical and surgical hospital, and the results were not generalized across settings.

The Mode of Therapy Used

Yes

Did the researchers use a standard mode of therapy? (e.g. psychoanalysis, rational emotive therapy, behavior therapy, etc.). Also, were those studies that used behavior therapy conducted over an extended period of

time (at least 5 sessions of therapy), and/or did they include a follow-up study?

The researchers used cognitive-behavioral treatment techniques. Patients in the treatment in couples condition group "received training in rational thinking techniques..." (p. 328-9).

Level of Pathology

Yes

Did the subjects have subclinical concerns? (e.g. anxiety, self-esteem, assertiveness, etc.) The study must have used patients with some level of psychological pathology, preferably patients with severe psychological pathology.

"Twenty-nine patients had low back pain, and 15 of these reported at least one additional type of pain... Patients reported having pain for an average of 16.5 years (SD = 12.6 years, range = 2-49 years)." (p. 327).

The Description of Therapy Used

Yes

Did the researchers specify a description of therapy or give its criteria of what was considered psychotherapy, and if so, was the



description specific enough that other researchers could replicate the therapy, if they wish?

The researchers provided enough information that other researchers would replicate the replicate the therapy. (p. 328).

The Description of Placebo Used

No

Did the researchers use a placebo group, and if so, did the study specify a description of placebo or give its criteria of what was considered a placebo? Also, was the description specific enough?

No placebo group was used. However, the researchers did include a waiting-list treatment group. The using a waiting-list as a control is not the best method for controlling for extraneous variables.

Due to the fact that the sample population was not a true representative sample of the unrestricted population of chronic pain sufferers, and the fact that the dependent measure was based on self-reporting, which is always questionable, and the fact that the study was not a true control evaluation, its findings are not only questionable, but the results cannot be

generalized to the unrestricted population of chronic pain sufferers.

Ordman, A. M., and Kirschenbaum, D. S. (1985).

CRITERIA: \_\_\_\_\_ YES NO

VALIDITY:

History

Yes

Were the subjects previously treated for the same symptoms, and did the previous treatment cause them to have a positive or a negative view of therapy?

Some subjects had been previously treated for bulimia and other related conditions. "Three subjects reported a previous history of Anorexia Nervosa, although 2 of them were never formally diagnosed or treated for it. Three of the clients had previously received treatment for bulimia, whereas 2 others had been in therapy for family and academic problems." (p. 306).

Maturation

No

Did the researchers allow more time to elapse before taking the final measure that they allowed between treatment applications? (e.g. if the subjects received treatment one time a week, did the researchers allow more than that one week to elapse before taking the measure

of the effectiveness of therapy?).

Measures were taken after treatment was terminated. Measures were also taken throughout treatment. (p. 306).

### Testing

Yes

1. Were the subjects tested repeatedly for the same effect or for different effects?

The subjects were tested at several times throughout the study.

Particularly, with this population, repeated testing is a problem.

Patients who think they may not be benefiting from treatment may respond falsely on the test to let the researcher believe that they are benefiting from treatment and vice versa.

2. Is the test validated (and cross-validated) for the types of subjects used in the study?

Yes

The researchers used several tests that appeared to be valid tests for bulimia. The primary test that directly measured

eating behavior was the Binge Questionnaire. It appeared to be a valid test for the purpose of measuring eating behavior. (p. 307).

Instrumentation

Yes

1. Was the testing procedure of the study consistent across subjects? (e.g. was every subject tested the same way and with the same test?)

All subjects were tested using the same assessment procedure. "Potential clients responded to a structured interview, providing answers to questions about their eating behavior and relevant demographic information." (p. 307). In addition, all subjects received the EAT as well as other tests.

2. Was the criteria of what constituted a therapeutic effect consistent across subjects?

Even though no criteria was specified out by the researchers, it appears that the achievement of better eating attitude and behavior, and improved

psychological adjustment was the measure of the effectiveness of therapy. (p. 308).

Selection Bias No

1. Did the subjects in both the experimental and control group have different symptoms?

All subjects in both treatment group had the same symptoms. (p. 306).

2. Was the subjects selected from different population? No

All subjects were selected from the University of Wisconsin Psychology Research and Training Clinic. (p. 306).

Randomization Yes

Were the subjects randomly assigned to the experimental and control groups?

"...and then were randomly assigned to either the brief-intervention-waiting-list condition (n=10) or the full-intervention condition (n=10)."  
(p. 306).

Reliability of the Measure

Yes?

Was the measure reliable? Did the instrument produce consistent results with repeated testing? (e.g. if several measures were taken before the treatment was administered, would each measure be the same?).

Even though the test may have been reliable, the measures were based on self-reported data. Particularly with this population, the reliability of test based on self-reporting, becomes increasingly questionable because these individuals have a distorted perception of their body image. Therefore, the accuracy of self evaluation by these individuals is questionable. (p. 307).

Interaction of History and Treatment

No?

Was the study conducted at a particular time and/or within a particular time period, where the characteristics of that time would negatively or positively affect treatment? (e.g. was the study conducted within a year after the end of the Vietnam War?).

There is a possibility the study may have been affected by this threat to external validity. The study was conducted in 1985 and the 80's era was an era when body image was a primary concern. This may have had an affect on the performance of the clients. Also, little was known about the disorder and people were not open about their disorder and rarely, openly sought treatment.

Interaction of Selection and Treatment

No

1. Were the results of the study incorrectly generalized to a different population than as used by the study?

"The results of the current investigation clearly indicate that clients who received the cognitive-behavioral treatment improved much more than those in the comparison-waiting-list group. (p. 310).

2. Did the researchers use "real patients," patients who were not solicited for participation in the study, and who were not "perfectly healthy" college students? Yes



Even though the subjects met the DSM-III criteria for bulimia, 18 of the 20 subjects were college students and all were solicited for participation in the study. (p. 306). Because the subjects were solicited for participation, they were likely to be individuals who believe that they would benefit from treatment.

Interaction of Setting and Treatment No

Was the treatment condition and the control condition conducted in the same setting, and were the results incorrectly generalized across settings?

The study was conducted at the University of Wisconsin Psychological Research & Training Clinic, and the results were not generalized across settings.

The Mode of Therapy Used Yes

Did the researchers use a standard mode of therapy? (e.g. psychoanalysis, rational emotive therapy, behavior therapy, etc.). Also, were those studies that used behavior therapy conducted over an extended period of time (at least 5 sessions of therapy), and/or

did they include a follow-up study?

The researchers used cognitive-behavior therapy and waiting-list condition. (p. 306).

Level of Pathology Yes

Did the subjects have subclinical concerns? (e.g. anxiety, self-esteem, assertiveness, etc.) The study must have used patients with some level of psychological pathology, preferably patients with severe psychological pathology.

All subjects met the DSM-III criteria for bulimia and had bulimia for 1 to 11 yrs (M=2.71, SD=1.8). During the 2 months prior to treatment, they reported vomiting 1.25 to 35 time per week (M=12, SD=5.7). (p. 306).

The Description of Therapy Used No

Did the researchers specify a description of therapy or give its criteria of what was considered psychotherapy, and if so, was the description specific enough that other researchers could replicate the therapy if they wish?

The researchers gave a description of the therapy used. However, it was a vague and incomplete

description of therapy. (p. 306-7).

The Description of Placebo Used No

Did the researchers use a placebo group, and if so, did the study specify a description of placebo or give its criteria of what was considered a placebo? Also, was the description specific enough?

No placebo group was used.

This study did not satisfy several of these standard criteria, most importantly, the primary dependent measure of the study was not reliable. Therefore, the reliability of the findings obtained is questionable. However, in addition to not satisfying other important criteria relevant to this study, researchers did not use a placebo or a control group. Without the use of a control of a placebo group, the results of the study could not be generalized to the unrestricted population.

Shea, et al. (1990).

CRITERIA: \_\_\_\_\_ YES \_\_\_\_\_ NO

VALIDITY:

History

Not Clear

Were the subjects previously treated for the same symptoms, and did the previous treatment cause them to have a positive or a negative view of therapy?

The researchers did not report information concerning the subject's past treatment history.

Maturation

No

Did the researchers allow more time to elapse before taking the final measure that they allowed between treatment applications? (e.g. if the subjects received treatment one time a week, did the researchers allow more than that one week to elapse before taking the measure of the effectiveness of therapy?).

"The patients were...assessed at termination of treatment covering several domains of outcome." (p. 712).

Testing

Yes

1. Were the subjects tested repeatedly for the same effect or for different effects?

"The patients were assessed before treatment, during treatment (4, 8, and 12 weeks), and at termination of treatment, on a battery of instruments covering several domain of outcome."  
(p. 712).

2. Is the test validated (and cross-validated) for the type of subjects used in the study?

Yes

The researchers used several test to measure several factors related to depression. However, the Hamilton Rating Scale of Depression was the primary measure of depression and it was valid.

Instrumentation

Yes

1. Was the testing procedure of the study consistent across subjects? (e.g. was every subject tested the same way and with the same test?)

"All of the patients were screened using

the same instruments. "Personality disorder were assessed by clinical evaluators at intake and treatment termination and by therapist following the second treatment session and at treatment termination." (p. 712).

2. Was the criteria of what constituted a therapeutic effect consistent across subjects? Yes

In addition to achieving a lower score on the other test used to measure the effectiveness of therapy, the achievement of a lower score on the Hamilton Rating Scale for Depression was the criteria for the effectiveness of therapy.

Selection Bias

No

1. Did the subjects in both the experimental and control group have different symptoms?
- All subjects met the Research Diagnostic Criteria (RDC) for current episode of definite major depression and had a minimum score of 14 on an amended version of the

Validity 130

17-item Hamilton Rating Scale for depression. (p. 712).

2. Was the subjects selected for different populations? No

It appeared that all subjects were selected for the patient population at the National Institute of Mental Health Center.

Randomization Yes

Were the subjects randomly assigned to the experimental and control groups?

"A total of 250 patients met study criteria and were randomly assigned to one of four treatment modalities..." p. 712.

Reliability of the Measure Yes

Was the measure reliable? Did the instrument produce consistent results with repeated testing (e.g. if several measures were taken before the treatment was administered, would each measure be the same?).

The researchers used several tests to evaluate different aspects of the sample population and all were reliable tests.

The Hamilton Rating Scale for Depression was used to evaluate depression, and it is a reliable test for assessing depression. (p. 712).

Interaction of History and Treatment

Not Clear

Was the study conducted at a particular time and/or within a particular time period, where the characteristics of that time would negatively or positively affect treatment? (e.g. was the study conducted within a year after the end of the Vietnam War?).

Do not apply.

Interaction of Selection and Treatment

No

1. Were the results of the study incorrectly generalized to a different population than as used by the study?

"It is important to emphasize the restrictions of the sample, particularly with regard to exclusion criteria for schizotypal features and antisocial personality disorder. Because of the selectivity of the sample, these findings cannot be generalized to the



unrestricted population of depressed individuals." (p. 713).

2. Did the researchers use "real patients," patients who were not solicited for participation in the study, and who were not "perfectly healthy" college students? Yes

"Subjects were male and female outpatients who met Research Diagnostic Criteria (RDC) (17) for a current episode of definite major depressive disorder and have a minimum score of 14 on an amended version of the 17-item Hamilton Rating Scale for Depression (18)." (p. 712).

Interaction of Setting and Treatment

No

Was the treatment condition and the control condition conducted in different settings, and were the results incorrectly generalized across settings?

The study was conducted at the National Institute of Mental Health Treatment Center.

The Mode of Therapy Used

Yes

Did the researchers use a standard mode of therapy? (e.g. psychoanalysis, rational emotive therapy, behavior therapy, etc.). Also, were those studies that used behavior therapy conducted over an extended period of time (at least 5 sessions of therapy), and/or did they include a follow-up study?

The researchers used cognitive-behavioral therapy interpersonal therapy, imipramine clinical management, and placebo plus clinical management. (p. 712).

Yes, the length of the study was 16 weeks, with 16-20 sessions." (p. 712).

Level of Pathology

Yes

Did the subjects have subclinical concerns? (e.g. anxiety, self-esteem, assertiveness, etc.) The study must have used patients with some level of psychological pathology, preferably patients with severe psychological pathology.

"All subjects met the Research Diagnostic Criteria (RDC) for a current episode of definite major depression disorder..."

(p. 712).

The Description of Therapy Used

Not clear

Did the researchers specify a description of therapy or give its criteria of what was considered psychotherapy. If so, was the description specific enough that other researchers could replicate the therapy, if they wish?

The researchers referred the reader to a larger study which reported the detailed description of the treatment procedure. (p. 712) [19]. However, it was not clear how close the treatment manual was followed.

The Description of Placebo Used

No

Did the researchers use a placebo group, and if so did the study specify a description of placebo or gave its criteria of what was considered a placebo? Also, was the description specific enough?

Yes, the researchers did use a placebo group, however the researchers did not report sufficient information on the placebo group to permit a valid

evaluation, they merely stated that a "placebo plus clinical management" group was used as one of the treatment modalities. (p. 712).

The researchers did not report information concerning several important aspects of the study. For example, the researchers did not report information regarding the subjects previous treatment history, they did not report information regarding how the sample population was obtained, etc. In addition, though the study was a controlled study, the researchers did not report sufficient information regarding the characteristics of the placebo group to permit a valid evaluation of the study.

Woody, et al. (1985).

CRITERIA: YES NO

VALIDITY:

History

Yes

Were the subjects previously treated for the same symptoms, and did the previous treatment cause them to have a positive or a negative view of therapy?

"The subjects had been receiving methadone treatment for at least two weeks but not more than six months during their current treatment episode..." (p. 1082).

Maturation

No

Did the researchers allow more time to elapse before taking the final measure that they allowed between treatment applications? (e.g. if the subjects received treatment one time a week, did the researchers allow more than that one week to elapse before taking the measure of the effectiveness of therapy?).

"The patients were tested at the start of treatment and at the one- and at the seven-month evaluation points." (p. 1082).

Testing

Yes

1. Were the subjects tested repeatedly for the same effect or for different effects?  
 "A series of self-reports psychological test measuring affect, cognition, and psychiatric symptoms was administered to the patients at start of treatment and at the one-and seven-month evaluation point." (p. 1082).

2. Is the test validated (and cross-validated) for the type of subjects used in the study?  
 Beck Depression Inventory  
 Maudsley Personality Inventory  
 Hopkins Symptom Checklist-90 items  
 Shipley Institute of living scale.

Yes

Instrumentation

Yes

1. Was the testing procedure of the study consistent across subjects? (e.g. was every subject tested the same way and with the same test?)  
 All subjects received the same test.  
 "A series of self-report psychological

tests measuring affect, cognition, and psychological symptoms was administered to the patients..." (p. 1082).

2. Was the criteria of what constituted a therapeutic effect consistent across subjects? Yes

See Table 1 and 2.

Selection Bias

No

1. Did the subjects in both the experimental and control group have different symptoms?

"Patients selected for the psychotherapy study were all men between 18 and 55 years of age, were nonpsychotic, did not have a persistent or clinically significant organic brain syndrome, and meet Food and Drug Administration requirements for methadone maintenance treatment."

(p. 1082).

2. Was the subjects drawn from the different populations? Not clear

The researches did not report data on how the sample population was obtained.

Randomization

Yes

Were the subjects randomly assigned to the experimental and control groups?

"Patients were randomly assigned to three treatment conditions on signing the consent form and completing intakes." (p. 1082).

Reliability of the Measure

Yes

Was the measure reliable? Did the instrument produce consistent results with repeated testing? (e.g. if several measures were taken before the treatment was administered would each measure be the same?).

"All test appeared to be reliable instruments. "The psychological tests are well standardized, have proven reliability and validity, and were administered under supervised conditions." (p. 1082).

Interaction of History and Treatment

No

Was the study conducted at a particular time and/or within a particular time period, where the characteristics of that time



would negatively or positively affect treatment? (e.g. was the study conducted within a year after the end of the Vietnam War?).

Does not apply.

Interaction of Selection and Treatment No

1. Were the results of the study incorrectly generalized to a different population than was used by the study?

"In the case of opiate-dependent patients, it does not appear beneficial to employ psychotherapy as a means of improving treatment outcome for those with Antisocial personality only."  
(p. 1082).

2. Did the researchers use "real patients," patients who were not solicited for participation in the study, and who were not "perfectly healthy" college students? Yes

"...Met Food and Drug Administration requirement for methadone maintenance treatment." (p. 1082). In addition, "subjects met DSM-III and RDC diagnostic

Validity 141

criteria." (p. 1082).

Interaction of Setting and Treatment No

Was the treatment condition and the control condition conducted in the same setting and were the results incorrectly generalized to a different setting?

It appeared that the study was conducted at the University of Pennsylvania.

The Mode of Therapy Used Yes

Did the researchers use a standard mode of therapy? (e.g. psychoanalysis, rational emotive therapy, behavior therapy, etc.). Also, were those studies that used behavior therapy conducted over an extended period of time (at least 5 sessions of therapy), and/or did they include a follow-up study?

"...Supportive-expressive psychotherapy plus counseling alone (SE) or Cognitive Behavior therapy plus counseling." (p. 1082).

Yes, therapy lasted more than 5 sessions.

Level of Pathology Yes

Did the subjects have subclinical concerns? (e.g. anxiety, self-esteem, assertiveness,

etc.) The study must have used patients with some level of psychological pathology, preferably patients with severe psychological pathology.

The subjects were previously diagnosed as having a personality disorder and they did meet the DSM-III and RDC criteria for that diagnosis.

The Description of Therapy Used

Yes

Did the researchers specify a description of therapy or give its criteria of what was considered psychotherapy. If so, was the description specific enough that other researchers could replicate the therapy, if they wish?

The researchers referred to a previous study where the therapy procedure was described.

The Description of Placebo Used

No

Did the researchers use a placebo group, and if so, did the study specify a description of placebo or give its criteria of what was considered a placebo? Also, was the description specific enough?

No placebo group was used. Also no control

group was used.

This study met several of the criterion of this thesis. However, because the subjects had received previous treatment, as early as two weeks before participating the study, it is likely that the study was affected by history effect. In addition, the patients were repeatedly tested through treatment, and the researchers did not use a placebo group to control for extraneous variables.