COMMON FACTORS IN PSYCHOTHERAPY

Essays in honor of Emeritus Prof. Dr. M.M. Nawas

Edited by Martin A. van Kalmthout Cas Schaap Franz L. Wojciechowski


No part of this book may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from the publisher.

Printed in The Netherlands ISBN 90 265 0639 2

CHAPTER 1 IN SEARCH OF PSYCHOLOGY
M.M. Nawas

Introductory
Toward a systematic psychology

CHAPTER 2 THE BASICS OF EXEMPLAR
M.A. van Kalmthout

Introductory
History
Extrapolation
Discussion

CHAPTER 3 CULTURE AND INTERCULTURAL STUDIES
P.W.M. Hofman

Introductory
Culture
Culture and communication
Problem-solving
The feedback loop

CHAPTER 4 NONSPECIFIC FACTORS IN HYPERTHERMAL RESEARCH
H.R.A. Beavers

Introductory
Nonpecific factors
Non specific factors in hyperthermia
Scientific aspects
A challenge
CONTENTS

PROLOGUE
M.A. van Kalmthout, C.P.D.R. Schaap, and F.L. Wojciechowski ................................... vii

CHAPTER 1 IN SEARCH OF THE NONSPECIFIC FACTORS IN PSYCHOTHERAPY: A SPECULATIVE ESSAY
M.M. Nawas, P.W.M. Pluk, and F.L. Wojciechowski .................1

Introduction 1
Toward specifying the nonspecific 6
Some research issues 26
Coda 32

CHAPTER 2 THE BASIC STRUCTURE OF PSYCHOTHERAPY AS EXEMPLIFIED IN HYPNOTHERAPY
M.A. van Kalmthout, and A.H.G.S. van der Ven ............43

Introduction 43
History 45
Extrapolation to psychotherapy in general 52
Discussion 55

CHAPTER 3 CULTURE AND HEALING: A STUDY OF COMPARATIVE AND INTERCULTURAL PSYCHOTHERAPY
P.W.M. Pluk, and F.L. Wojciechowski .........................59

Introduction 59
Culture-boundedness of psychotherapy 59
Culture change and healing 60
Communalities and divergencies 62
Intercultural psychotherapy 64
Problems and pitfalls of intercultural psychotherapy 65
The feasibility of intercultural psychotherapy 70

CHAPTER 4 NONSPECIFIC FACTORS IN THE TREATMENT OF ESSENTIAL HYPERTENSION: A CHALLENGE FOR BEHAVIOR THEORY, RESEARCH, AND THERAPY
H R.A. De Mey ............................................77

Introduction 77
Nonspecific factors defined 78
Nonspecific factors and the treatment of essential hypertension: A survey 79
Scientific explanation 82
Nonspecific factors and scientific explanation 86
A challenge for behavior theory, research, and therapy 88
Seven chapters (2 through 7) have been expressly written for the occasion of his retirement, and the tribute is paid for this now belongs to The world over, in the field of practice, he has been keenest enthusiasts for Eysenck's unmitigated blessing. The present hazards. We have di\ed, among others, the "holier-than-thou" attitude of with the missionary energies in pursuit of the utilization to broaden the horizons and increase the effectiveness this now belongs to the.

History has many is the story of Hans Eysenck's SS if he were land of birth, Germany wanted to study psychology entry because he had alternative, his choice have had him!

Eysenck's contribu- therapy is concerned with the small packages" proposed, not free of bias, tesifications was not of
Seven chapters (2 through 8) of Common factors in psychotherapy have been expressly written in honor of Prof. Dr. M.M. Nawas on the occasion of his retirement. A delineation of why the honor was earned and the tribute is paid follows later.

The world over, psychotherapy both as a research enterprise and a field of practice has grown rapidly, perhaps more rapidly than its keenest enthusiasts had thought possible. A too rapid growth is not an unmitigated blessing. Fragmentation and loss of direction are ever-present hazards. We know from the past that these hazards had resulted, among others, in "schools" of psychotherapy each protecting its turf with ferocity; in closed psychotherapeutic societies each holding a "holier-than-thou" attitude shouting "eureka" and spreading its "word" with the missionary zeal of deluded religious cults; in squandering energies in pursuit of the "Grand Prix" which should have been mobilized to broaden the base of the budding science of psychotherapy and increase the effectiveness of its applied wing. Fortunately, much of all this now belongs to the annals of the past.

History has many twists of which some are fortunate. One such twist is the story of Hans Jurgen Eysenck. Realizing that he had to join Hitler's SS if he were to stand a chance of entering a university in his land of birth, Germany, Eysenck left at the age of 18 for England. He wanted to study physics at London University. He was not allowed entry because he had not taken the "right" prerequisite courses. As an alternative, his choice fell on psychology. How lucky is psychology to have had him!

Eysenck's contributions to psychology are many. As far as psychotherapy is concerned, our gratitude goes foremost to a mere six-page article he published in 1952; the saying that "precious gifts come in small packages" proved in this case to be true. This milestone, though not free of bias, touched off a chain reaction. Affected by its reverberations was not only the field of research in psychotherapy but
equally also those who for long held a strangle grip on the fortresses of power -- the dogmatists in the politics of practice and training whose power is now hopefully sapped for the good of all.

Contemporaneous with the shockwaves caused by Eysenck's report was another relevant development also of crisis proportions underway: The attack on the "medical model". This issue was soon expanded from its original intent to encompass questioning the value of one "expert" (psychiatrist) over another "expert" (psychologist), and thereafter to questioning the value of "expertise" as such -- for so long taken for granted.

A crisis is often therapeutic. The impact of a double dose cannot be but profound. When the dust settled down, constructive questions began to be raised in an admirably detached and sober sphere around the dual and intertwined issues of effectiveness of psychotherapy as a whole and of the values of training of whatever mold or length. The first issue came finally to revolve around the "magical" two-thirds improvement rate which seemed to cut across modalities, academically-anchored therapies and "alternative" and even "primitive" healing practices. And since the publication by Margaret Rioch of her well-designed study -- in which both psychology and psychiatry have cooperated -- the question began nagging about the logic, manner, direction, and wherefore of therapeutic training and specializations.

The issues raised as a result of the findings on both outcome and training will continue to engage attention. Whatever more refined research will in the future unfold, it cannot be denied that psychotherapy was humbled -- which is a good antidote against arrogance. The psychotherapeutic community began to take a hard look at itself and then a palpable unity of purpose emerged. Especially in the decades of the sixties and seventies concerted efforts have come to focus on identifying the factors which may well be common to (a) all forms of psychotherapeutic intervention, and (b) to the activities of those who practice it no matter by what name they may be called. We believe that the entire psychotherapeutic community is in debt to Jerome Frank, in particular, for lending his talent and well-earned reputation and prestige for encouraging this search which now assumes the status of a conviction which we share. Arthur K. Shapiro's work on the placebo phenomenon earns no less credit.

Common factors in psychotherapy, quite appropriately, opens with a chapter whose senior author is the man in whose honor the book is written. The chapter bosom lies the conviction that contents is detailed, subsequent chapters.

Chapter 2 (Van Kampen) "case-study". In the phenomena, they actually which chapter 1 peer to the same terrain mic closer reconnaissances therapies.

Chapter 3 (Pluck & Rich) which permit a broad of the Western mold have gleaned to the a front dealing with clinical their own.

Chapter 4 (De Mey) his attention also on draws some conclusion -- with a special mess is one -- to come to g therapy as he sees them.

Chapter 5 (Broekk) the latest and what is relatively pitfalls-safe research. This chapter whether the so-called course, their intent is

Chapter 6 (Wojciechowski) in that both points -- and further also affinity to that other flavor. Wojciechowski systematic though brief points of the past. He could describes how he implanted obtained.

The line of continuity from past failures of
written. The chapter sketches in broad strokes the framework in whose bosom lies the conviction of the contributors. Inasmuch as the table of contents is detailed, no more than a few words need to be said about subsequent chapters.

Chapter 2 (Van Kalmthout & Van der Ven) treats hypnotherapy as a "case-study". In the authors brief historical scrutiny of this phenomenon, they actually substitute the macrocosm-oriented lenses through which chapter 1 peered, by another set of optic lenses designed to scan the same terrain microcosmically, thereby narrowing the focus for a closer reconnaissance of what they call the "basic structure" of the psychotherapies.

Chapter 3 (Pluk & Wojciechowski) reverts back to using the lenses which permit a broad sweep. They take the reader beyond the confines of the Western mold but then trace the path back to relate what they have gleaned to the realities which therapists in the West have to confront dealing with clients whose (sub)cultural backgrounds differ from their own.

Chapter 4 (De Mey) bears affinity to chapter 2. Its author focuses his attention also on a circumscribed topic -- hypertension. De Mey draws some conclusions, then challenges all contributors to this volume -- with a special message addressed to behavior therapists, of whom he is one -- to come to grips with the realities of science and its philosophy as he sees them.

Chapter 5 (Broekman, Schaap, & Lange) acquaints the readers with the latest and what its authors consider to be the most promising and relatively pitfalls-safe paths for designing and implementing outcome research. This chapter raises, among others, the question as to whether the so-called "common factors" are real or simply artefacts. Of course, their intent is to bridle enthusiasm, not to extinguish it.

Chapter 6 (Wojciechowski) is in a real sense supplementary to chapter 5 in that both deal with outcome -- albeit from different vantage points -- and further raise the issue of artefacts. This chapter bears also affinity to that of Van Kalmthout and Van der Ven in its historical flavor. Wojciechowski takes the reader, as a historian does, into a systematic though brief purview of the vicissitudes (and sometimes foibles) of the past. He concludes by presenting the design he favors, describes how he implemented it, and touches on some of the results he obtained.

The line of continuity between the outcome-oriented contents of chapters 5 and 6 is maintained in chapter 7 (Schaap). Its author draws from past failures of psychotherapeutic research the lesson which dic-


states the need for a focal attack on the issue of the processes involved in interactions. He illustrates his presentation by describing various options and, of course, he, too, has his own "pet".

Of all the contributions to Common factors in psychotherapy, chapter 8 (Donker & Schaap) is least directly tied to the title of the book. Indirectly, the bond is of course present -- as is illustrated in the text of this Prologue. The chapter sketches the trials and tribulations of the growth of psychotherapy as an independent discipline, identifies some real dilemmas, and closes by acquainting the reader with the regulatory plans currently on the drawing board in the Netherlands.

The book closes with a short Epilogue written by Emeritus Prof. Dr. Nawas.

And now it is time to follow up on the opening of this Prologue. As editors, we consider it a privilege to mark the occasion of Prof. Dr. Nawas’ retirement by touching on only two challenging highlights of his contributions to clinical psychology and psychotherapy. The first challenge goes back to the last two years of his life in the U.S.A. The second highlight-spot is associated with his acceptance of our University’s invitation to him in 1972 for moving to the Netherlands to head the Department of Clinical Psychology.

In 1970, Prof. Dr. Nawas left the University of Missouri for Indiana with the intent of developing training programs in the clinical field which some have called "very radical" but of which several have since been implemented in many universities. Upon his departure from Missouri, the 14 men and women whose dissertations he was then supervising presented him with a testimonial which we have contrived to get hold of. The testimonial speaks for itself and for us. It reads in part:

IN APPRECIATION

These words are meant to express, however inadequately, our grateful appreciation for the years of help and encouragement we have received from Prof. M.M. Nawas, who has sometimes heroically and always unstintingly worked with the interest of his students and colleagues as one of his most sacred trusts. He has taught us the force and life in the ideals of integrity, charity, and excellence. Our gratitude is exceeded only by our affection for him.

Our University’s invitation to Prof. Dr. Nawas to come to our country was based on a background of distinction. This background includes his completion of his studies at the University of Chicago in record time (and, due to his excellent F. 100,000 it then could be promoted from instructor to professor (all within 3 years), publications on diverse ones have been cited in his field and prominence and election as "fellow" by numerous recognitions. It sound like a eulogy.

Emeritus Prof. Dr. Nawas’ roots from the U.S. immense courage and some of the culture shock of three children (then " found trust in the ad in general would take a family is concerned, this

Restricting ourselves was needed in our de course in this field was remedy this glaring psychotherapy as he thought were not clinical psy clinicians and highly who have to do with ph completed their trainin status as well. Thanks efforts, inspiration, an throb with vitality. Th proved optimal for the in psychotherapy”) of in the variegated contr some, not all, of the our department and whi
(and, due to his excellence from start to finish, a waiver of the nearly F. 100,000 it then cost other students); the speed with which he was promoted from instructor to assistant professor to associate professor to professor (all within 7 years, very unusual for those days); his many publications on divergent aspects of psychology -- of which relevant, ones have been cited by various contributors to this volume -- notably those in the field of behavior therapy which he helped bring into adulthood and prominence without abandoning his eclectic outlook; and, his election as "fellow" by the American Psychological Association -- one of numerous recognitions we need not mention lest our tribute begins to sound like a eulogy. Only a few more words and then we stop.

Emeritus Prof. Dr. Nawas was at the age of 47 when he pulled all roots from the U.S. and brought his family along. Only a man with immense courage and commitment would risk, among others, the impact of the culture shock on his wife and himself at their age, and on their three children (then 19, 12, and 8). And only one with equally profound trust in the adaptability, resourcefulness and stamina of people in general would take such a daring step. At least in so far as his family is concerned, this trust has been fully vindicated.

Restricting ourselves to only one aspect -- psychotherapy -- of what was needed in our department in 1972, suffice to say that not a single course in this field was taught and research in it was nonexistent. To remedy this glaring deficit and, at once, to broaden the base of psychotherapy as he thought it ought to be, Prof. Dr. Nawas expressly sought some staff members -- of whom the senior editor is one -- who were not clinical psychologists. These have since become trained as clinicians and highly committed ones; all the staff in our department who have to do with psychotherapy (by far the majority) have also now completed their training as psychotherapists and have supervisory status as well. Thanks largely to Prof. Dr. Nawas' vision, sustained efforts, inspiration, and the climate of freedom he created for all members of the staff to pursue their individualized, self-chosen research directions and preferred therapeutic orientations, our department now throbs with vitality. The interdisciplinary blend, nurtured by freedom, proved optimal for the growth of the seed which he had planted upon his arrival and which has now grown into a tree (the "common factors in psychotherapy") of several branches. These branches are reflected in the variegated content of this volume -- and the content represents some, not all, of the research activities which are flourishing in our department and which are having a contagious effect on our students.
It is time to close. Thanks to the many who have helped us in making the completion of this volume on time possible, particularly Carmen Jansen-Nawas. To the well-being of Emeritus Prof. Dr. Nawas and his family go our heartfelt wishes.

Nijmegen, March 26, 1985

Department of Clinical Psychology
Catholic University Nijmegen

1. INTRODUCTION

Any form of intervention; brainwashing does not go; some yet insufficiently understood conditions
EVALUATING THE EFFECTS OF PSYCHOTHERAPIES:
COMMON FACTORS OR ARTEFACTS?

T.G. BROEKMAN
C.P.D.R. SCHAAP
M.M. LANGE

1. INTRODUCTION

Two thirds of the clients in psychotherapy generally improve. The different psychotherapeutic schools have about the same improvement rates (Luborsky, Singer, & Luborsky, 1975). These are two of the conclusions that are often drawn from the body of results of psychotherapy outcome research. Moreover, they are often interpreted as being arguments in favor of common factors in psychotherapy (Garfield, 1980). What we plan to do in this essay is to have a critical look at these "facts" from a methodological perspective; for we think that these so-called "common factors" may for a part be artefacts and misinterpretations, caused by flaws in the design and measurement of outcome studies all operating in the same direction (Kazdin, 1983). We will focus on methodological issues and, therefore, refrain from presenting the umpteenth review of the results of psychotherapy outcome research. This has been done comprehensively and well by others (e.g., Bergin, 1971; Meltzhoff & Kornreich, 1970; Rachman & Wilson, 1980).

Let us start off by some remarks about the latest way of integrating research results, meta-analysis. Because it is a formalized procedure, meta-analysis enables us to set the stage handsomely for the five problems we will discuss in the next sections:

1. Difficulties that are associated with the assessment of the complaints of clients, which in turn influence the dependent variables in the outcome study -- the effect criteria.
2. Related to this is the measurement of these independent variables and in particular the measurement of change. We will argue that part of the common results in psychotherapy outcome research could be due to factors associated with the measurement of the outcome variables.
(3) On an abstract level, the factors that are linked to measurement also account for the tenuous relationship between the therapy itself (the independent variable) and its effects. Incomplete specification of treatment, together with the assumption that therapy will be carried out as planned, are the major causes.

(4) A special problem in outcome research is the implementation of the design of the study. Difficulties in implementing a design that is logically compelling, have consequences for the selection of clients and therapists in a direction that may well obscure differential results.

(5) An important theme in the interpretation of the results of psychotherapy outcome research is the blurring of the functions to which the results will be put to use. For instance, research designed to test explicit hypotheses derived from theory, generally cannot assist administrators or practitioners in optimizing psychological health care. It is particularly with this in mind that we will close this essay by recommendations for future research.

2. META-ANALYSIS

In this section on meta-analysis we want to illustrate five aspects of psychotherapy research that contribute to the lack of differential results, namely, criteria, measurement, treatment, the quality of the studies, and the role of the studies. Meta-analysis is particularly well suited for this purpose because it is a quantitative technique and therefore forces researchers to make explicit their decisions in manipulating the data. In a narrative review most of these decisions are implicit and their detection by the reader is not at all easy.

Meta-analysis has received much attention and already pervades the journals as a new and promising technique for integrating and reviewing particular fields of research (e.g., Wampler, 1982; Nicholson & Berman, 1983). To date several meta-analyses have been reported in the field of psychotherapy research (Andrews & Harvey, 1981; Landman & Dawes, 1982; Prioleau, Murdock, & Brody, 1983; Shapiro & Shapiro, 1982; Smith, Glass, & Miller, 1980). In this technique the results of each dependent variable in each outcome study are expressed as the differences between the means of the treated and control groups divided by the standard deviation of the control group. These differences, called effect-sizes (ES), are then averaged within studies (Landman & Dawes, 1982; Prioleau et al., 1983) or each of them is treated as a new independent variable (Glass & Kliegl, 1983; Shapiro & Shapiro, 1982,
1983; Smith et al., 1980). The impact on effect size of several characteristics of studies (such as treatment method, client and therapist variables, and measurement and design features) is determined empirically. Smith et al. (1980) contrasted this method with the traditional box-score method (Bergin, 1971; Eysenck, 1966; Luborsky et al., 1975; Meltzhoff & Kornreich, 1970; Rachman & Wilson, 1980), in which results are described narratively and tabulated in a summary table consisting of a number of significant studies broken down by other variables of interest.

Criteria. The technique calls for a decision regarding the unit of analysis. Landman and Dawes (1982) averaged the different effect-sizes within one study, thus treating the study as the unit, whereas Smith et al. (1980) treated all individual criterion measures as independent data. It is obvious that without weighting, which Smith et al. did not do, those studies with relatively more criteria are automatically overweighted. On the other hand, when construing the study as the unit it is implicitly assumed that the different criteria are equally significant (Spence, 1983) -- a very dubious assumption, since it is rather obvious that some criteria are more relevant than others. Cordray and Bootzin (1983) pointed out a clear example from a study by Warner (1969), which used four outcome measures in studying group therapy for alienated students. While the average effect-size was .12 for the four measures, the effect-size for "alienation" was .56, this being the outcome measure which one would expect to be the most relevant to this particular complaint and population. In section 3 we will discuss the issue of the importance and relevance of criteria further.

Measurement. Another important aspect of outcome is the way it is measured. In an attempt to determine the influence of the method of measurement, Shapiro and Shapiro (1982) coded the outcome measures on (1) domain, (2) specificity, (3) reactivity, and (4) technology. They found positive correlations between these four aspects of the method of measurement and effect-size. A more tractable domain, and a more specific, and more reactive measure, generally softer in technology, was associated with a larger effect-size. Unfortunately, it is not possible in meta-analysis to estimate the degree to which the method of measurement contributes to the variance of the measure itself. It is only possible to categorize the measure and to correlate it with the effect-size. In section 4 we will elaborate on this issue.
**Treatment.** To be able to analyze the influence of the characteristics of the individual studies on effect-size, a coding scheme has to be devised. Since consensus on such a categorization is lacking, particularly in so far as the treatment modality is concerned -- Smith et al., for example, consider Rational Emotive Therapy a verbal psychotherapy and Meichenbaum's Self-instructional Training a behavioral therapy -- it is hard to compare the results of different meta-analyses. We will return to the classification of the treatments in section 5 where we discuss the definition and the implementation of the treatment.

**The design of the studies.** Another important issue is the quality of the individual studies which are selected as the data base for the meta-analysis. Meta-analysis is as good as the data it is based on (Eagle, 1983). One of the assumptions of meta-analysis is that different problems in the quality of the individual studies cancel each other out. Consistent findings are therefore robust and veridical. However, as Kazdin (1983) states, this rationale can be challenged since many outcome studies suffer from similar problems or different problems that operate in the same direction, i.e., showing little treatment effect -- a problem succinctly summarized by Eysenck (1978) as: "garbage in, garbage out". Kazdin therefore advocates meta-analyses of those studies which were shown in advance to meet the highest standards of scientific rigor (and clinical relevance). However, this may be difficult to accomplish. It is clear that the individual studies in the meta-analyses are unrepresentative of clinical practice, as will be argued in section 6, and more relevant studies are needed. But methodological rigor and clinical relevance might be incompatible, as is suggested by a number of studies (Nawas & Pucel, 1971; Emmelkamp & Walta, 1978; Wojciechowski, 1984). An increase in methodological rigor will necessarily impart to subjects a feeling of partaking in an experiment instead of a therapy -- thereby automatically causing a decrease in therapeutic effect.

**Role.** We fully agree with Robert Rosenthal (1983) who emphasizes that, since a meta-analysis is made up of a series of judgments, it is as much in need of replication as are the individual studies that make up the meta-analysis. At the same time it is important to make the decisions and assumptions in each analysis as explicit as possible. Much of the criticism from both the opponents and proponents of meta-analysis in the field of psychotherapy, is directed towards the heterogeneity of the studies sampled (Eagle, 1983; Fish, 1983; Maher, 1983; Shapiro, 1983). As Maher states, who would try to answer the question "Do
doctors cure illness?". One would want to know with some precision which doctors doing what to which kind of sickness -- a criticism which echoes Kiesler's (1966) identification of the "uniformity myth". The problem of the heterogeneity is further complicated, as Mintz (1983) notes, by the confounding of variables such as type of treatment and disorder (for example, cognitive-behavioral and depression, respectively).

We think that the technique of meta-analysis holds much promise, only, however, when the problems we will discuss in the following sections are dealt with in the individual outcome studies.

3. ASSESSMENT AND CRITERIA

On an abstract level there is no lack of clarity about the aim of psychotherapy: effectuating change in a client(system). Studies on the effect of psychotherapy should focus on these changes.

Change implies that when initial assessment is carried out at time t1, there exists a condition A and that at a following time (t2) there exists another condition, B. A client(system) can obviously be characterized by a number of states (A1.....An). A1 may be a certain amount of hormonal secretion, A2 a particular sense of well-being, A3 a characteristic personality trait, A4 a certain psychological complaint, A5 a particular behavior pattern, and so on. To be able to study the effect of psychotherapy the researcher has to be selective. Not all states A1.....An are relevant, and one simply cannot study or monitor them all.

Effect criteria have to be formulated as well. These may differ widely with regard to stability (personality versus blood pressure), observability (communication versus well-being), relatedness to a theory (ego-strength versus type-token ratio), changeability (rigidity versus relaxation skills), and level of organization (physiological, emotional, behavioral, cognitive); all of these being factors that might influence the reported results. Moreover, in each of these criteria changes may be detected, characterized by direction, magnitude, time, importance, and source (Meltzhoff & Blumenthal, 1966).

Because change can take place on any number of dimensions and because change can be interpreted from a number of different perspectives or sources, most literature surveys advocate the use of more than one criterion. Strupp en Hadley (1977) "developed" the so-called tripartite model which is in reality a name they gave to the very same sug-
gestions made much earlier by Gurin, Veroff, and Feld (1966). The distinction is being made between three independent points of view, viz., the consumer's (client/system), the professional's, and the point of view of society. Cromwell, Olson, and Fournier (1976) discuss the same issue in the context of marital and family therapy. Society considers the therapy a success when the client is able to behave in an orderly fashion and according to the rules of society. Client's improvement is indexed by a sense of well-being or a decrease in complaints. Finally, the professional views improvement of a client in terms of a more or less explicit theoretical framework (e.g., personality restructuring, self-esteem, dynamic processes). Succinctly stated, society emphasizes the behavior of the client, the client the personally felt effect (including its affective aspects), and the professional the psychological structure of the client. Instead of talking about the effect of psychotherapy, each study will have to take this configuration into account, and speak of the effects of psychotherapy. In the words of Strupp and Hadley (1977), "A truly adequate comprehensive picture of an individual's mental health is possible only if the three facets of the tripartite model of functioning -- behavior, affect, and inferred psychological structure -- are evaluated and integrated" (p. 196). However, it will not do when these different orientations of values are reflected within the measures that are used to evaluate psychotherapy. For, when the data themselves are to a great extent determined by judgments based on different value orientations, they will be confounded. Moreover, they will be confounded by processes which are worthy of study in their own right.

Although the importance of using multiple criteria has been stressed, the criteria for evaluating psychotherapy were, until recently, limited to those that assessed the amount of client change. Besides the issue of client change (efficacy), we note an increased emphasis (e.g., Kazdin & Wilson, 1978) on additional criteria such as efficiency, costs of administering treatment, and financial and emotional costs to clients.

An element often overlooked in outcome is the client's level of functioning at the end of therapy. Garfield (1981) notes that outcome research has only investigated change without paying much attention to a satisfactory level of client functioning. He notes that many studies indicate that after therapy clients function at a level that is in no way comparable to that of "normal" subjects.

There are two final points which make this issue even more complex. One is the fact that psychotherapy may have negative effects. A sociophobic client may as a result of an assertiveness-training go out dating
to such an extent that his/her study progress will be impeded, very much to the distress of parents and teachers. So the different perspectives may clash.

The second point is that in many studies -- particularly in the field of behavior therapy -- a simple criterion or complaint (e.g., agoraphobia) is simply noted. Usually, no detailed information is given, nor has an effort even been made to gather information about the context of the complaint. In short: a behavioral analysis is lacking. One client may have become agoraphobic after a traumatic event; for another client the complaint may have a clear function in a marital relationship, etc. In the first case, an isolated complaint, the treatment indicated may be systematic desensitization or cognitive restructuring, whereas in the second case it may be marital therapy. This difference in context will, amongst other things, have consequences for the criteria to assess change. Essentially, we are referring here to the distinction between target complaints and therapeutic goals -- an issue dealt with elsewhere by Broekman & Schaap (1984). Moreover, clients who are referred for psychotherapy usually have more than one complaint: a phobia plus headache plus marital problems plus subassertiveness etc.; though one may be more central than the others, they all, nonetheless, should be reflected in the criteria selected for evaluating change. Even if the agoraphobia is treated as the criterion (and rightly so) one should also indicate the improvement with respect to those other factors, such as marital functioning.

4. PROBLEMS OF MEASUREMENT

The aforementioned criteria for evaluating change have to be assessed using a particular method or combination of methods. In this section, it is argued that psychotherapy outcome measures, particularly measures of change, share a good deal of variance totally unrelated to the concepts they are supposed to measure. It is particularly this phenomenon that might account for the relative lack of differences in efficacy between different treatments.

4.1. The method factor

In 1959 Campbell and Fiske already discussed the validation of a measuring instrument in terms of convergent and discriminant validity. They
argued that not only should the convergent validity of a new test be assessed, but the discriminant validity as well: a new test should be considered invalid if it correlates highly with another test from which it should differ. Because every test score is determined not only by the trait it is supposed to measure but by the method of measurement as well, estimates of the contributions of the trait and the method to the test score should be made. To assess the discriminant validity and the amount of trait-variance and method-variance, several traits have to be measured using several methods. The \( t \times m \) correlation-matrix of \( t \) traits measured by \( m \) methods, is called the multitrait-multimethod matrix. In this matrix the correlations between the same trait, measured by different methods, have to be higher than the correlations of these methods when assessing different traits. If not, the test score is, therefore, determined more by the method of measurement than by the trait it is assumed to measure. After discussing some examples reported in the literature, Campbell and Fiske conclude that "the typical case shows an excessive amount of method variance, which usually exceeds the amount of trait variance" (p. 93).

Cartwright, Kirtner, and Fiske (1963) analyzed outcome data -- from a client-centered psychotherapy project -- with the aim of finding out if the personality changes concomitant with the therapy could be described by one single general factor or by a multiple-factor representation. They used data (which were raw change scores as well as improvement ratings) from the client, the therapist, and a diagnostician. The results of the factor-analyses indicated that the data could be described by four factors which, however, did not reflect any theoretical concept that had been used in the study. Instead, the factors were associated with particular combinations of observer and measuring instrument, for which the authors used the term method factor. With regard to the impact of method factors, they noted that "the presence of method factors seriously impedes any attempt to determine the specific substantive changes occurring in any given client" (p. 174).

4.2. Method factor and method of analysis

The presence of method factors in the data of Cartwright et al. (1963) has been criticized by Mintz, Luborsky, and Cristoph (1979). They re-analyzed these data by using two-step factor-analysis: first extracting components within each source, building composite scores for the source, and then extracting components between the sources.
Moreover, they did not use raw difference scores, but residual gain scores -- as suggested by Manning and Dubois (1962). They concluded that there was substantial agreement about treatment outcomes amongst the different sources (client, therapist, diagnostician). In the same vein, Berzins, Bednar, and Severy (1975) stated that the presence of method factors in the Cartwright et al. results is mainly caused by the fact that the data were analyzed by means of a simple factor-analysis, thus capitalizing on within-instrument and within-source variance. Therefore, they advised the use of multivariate techniques, such as canonical analysis. In a study of 79 clients receiving psychotherapy, Berzins et al. (1975) applied these techniques and reported considerable agreement between the client and a diagnostician on dimensions of change such as neuroticism and psychoticism. Moreover, the measures of the therapist and the client clearly reflected consensus on the concept of self-acceptance. Consequently, it is clear that the finding of a method factor depends on the method of analysis. The solutions offered by Mintz et al. (1979) and Berzins et al. (1975) seem to be an elegant way of eliminating method factors, but their strategy is a wrong one for two reasons.

First, instead of extracting only common variance from the variables, one could specify the contributions of the different methods to the measurement of a concept. As Jöreskog (1974) shows in a re-analysis of the multitrait-multimethod matrix used by Campbell and Fiske (1959) that the matrix can be analyzed by explicitly assuming the presence of one or more method factors. Five traits were measured by three methods (self-ratings, teammate-ratings and staff-ratings). The hypothesis that the data could be represented by five trait factors was rejected. The final model, which fitted the data well, was one with five trait factors and two method factors (one for self-ratings and staff-ratings, and one for teammate-ratings). The result of the analysis gives, for each combination of trait x method (e.g., assertiveness rated by staff), separate variance components for trait, method and measurement error. This way one obtains an estimate of which method best measures a given trait.

Second, one should try to substantiate a method factor. Labeling part of the variance "method factor" does not explain why the person is responding the way he does. Werts, Jöreskog, and Linn (1972) speak of "the variety of test response tendencies covered by the rubric 'method factors'" (p. 676). Good examples of the substantiation of a method factor are demand characteristics (Orne, 1962) and other social factors potentially affecting the performance of a subject during, say, a
Behavioral Avoidance Test. As has been amply demonstrated (e.g., Nawas, 1972), subjects participating in a Behavioral Avoidance Test who receive a neutral instruction perform differently from subjects who receive the instruction that the test is meant to determine the effectiveness of a therapy. As Bernstein (1973) noted: "If situational variables are consistently found to be important in avoidance test settings, the discriminant validity of a paper-and-pencil test may be found to covary with them. If this is the case researchers interested in choosing a verbal report measure will need data on its situational validity" (p. 247). Stated differently, the subject is responding to the assessment device under control of the specific situation; this provides a substantiation of the method factor of the test involved.

The lack of specificity of data is particularly of telling import in judgmental data on which psychotherapy outcome research depends so heavily. For instance, Shapiro and Shapiro (1983) found in their database that 41% of the outcome measures were psychometric instruments and 27% were self-ratings, rendering a total of 68% of the data judgmental. As Fiske (1974) notes: "Most of the data are the products of complex interpretive judgmental processes within observers; the agreement between sets of observations is limited; the ties between observations and concepts are tenuous and inadequate" (p. 3).

The analysis of method factors can be extended to the study of change. The most commonly used procedures in psychotherapy outcome research involve the creation of a new variable: the raw difference between pre- and postmeasures or the residual gain score (Manning & DuBois 1962). As has been amply demonstrated (Harris, 1963; Cronbach & Furby, 1970; Rogosa, Brandt, & Zimowski, 1982), these procedures are inadequate for a number of reasons. Instead, researchers should pursue the model wherein the pretest data are treated, amongst others, as a predictor of the final status. Furthermore, Werts et al. (1972) show that within structural equation analysis (Jöreskog, 1970) the researcher can use the multitrait-multimethod model for the study of change, thereby explicitly accounting for the method factors.

It is clear that psychotherapy outcome measures (especially measures of change) share a good deal of variance that has nothing to do with what these measures were supposed to measure (Broekman, 1984). This is one of the reasons why we should not be surprised to encounter no remarkable differences in effects of different therapies in the literature. In the next section we will discuss another reason.
5. TREATMENT

Treatment, the independent variable in the study of outcome, is a major source of the communality of results in outcome research. This is due to the fact that a considerable amount of overlap exists between the different treatment conditions that are being compared. The degree of similarity in treatments obviously varies from one study to another; still it must be considered as being substantial. To illustrate this we will limit ourselves to three studies.

Hill, Thames, and Rardin (1979) analyzed psychotherapy sessions conducted by Rogers, Perls, and Ellis with the one and the same client. They concluded that the behavior of these three therapists was indeed very consistent with their theoretical "stipulations". However, if one chooses to go beyond the surface of differences and stress the communality instead, one notices that these divergent psychotherapists emitted remarkably similar behaviors. Eymael (dissertation in progress), in a study comparing the effectiveness of client-centered therapy and behavior therapy, reports a remarkable amount of overlap in the techniques between these two treatments. Finally, in an interview (Knoppers, 1980), one of the authors of the Temple study, Staples (Sloane et al., 1975), was confronted with the fact that in their outcome study, the techniques which were used by the behavior therapists accounted, post-hoc, for only about 50% of the time spent during psychotherapy. The question was: What did they do during the rest of the time? His answer was something like: what all therapists do, such as warming-up the clients -- a tactic which does not fall strictly under the banner of behavior therapy, at least as advocated in the 1960s and early 1970s.

In an attempt to define behavior therapy, Erwin (1978) refers to the concept of family resemblance, developed by Wittgenstein. It may be that a game of one sort possesses certain important characteristics A, B, C; another game may resemble the first in having properties B and C, but not A; and a third game may possess properties C and D, but not B. It is our view that the concept of family resemblance rightly refers to both the properties that all therapeutic treatment share with each other, the common factors in the treatments, and the properties they do not share.

Usually, treatment proceeds without prior specification of the independent variable. This is especially the case when the complaints do not lend themselves to treatment by simple and clear procedures. Consequently, the individual therapist exercises a considerable degree of freedom in conducting the therapy. However, careful researchers do
try to restrict this freedom. For example, Sloane et al. (1975) first offered the therapists in the two treatment conditions (behavioral and psychodynamic) a list of properties making up a stipulative definition of the treatment; only when the therapist agreed with this definition, he was included in the study.

Another valuable tactic is the utilization of written manuals and the training of therapists to use them such that "it is possible to feel relatively confident that therapists who claim to be using the same method actually are doing so" (Frank, 1979, p. 312). Luborsky et al. (1982) studied the degree to which independent judges could recognize three different manual-guided psychotherapies (drug counseling, supportive-expressive and cognitive-behavioral). After a considerable amount of therapist training, the judges were able to recognize what the intended form of therapy was from tape recordings in 70% to 80% of the sessions. The authors commented: "This level of performance can be considered anywhere from reasonably adequate to very successful, depending on what one expects" (p. 59). Our expectation is that, in the modal outcome study, judges would perform far less well (than the 70-80% figure), not necessarily because of the qualities of the judges but as a consequence of the lack of specificity of the treatment and training of the therapists.

A third tactic calls for monitoring the actual delivery of the treatment. A measure should be constructed to permit a quantifiable statement about the degree to which the actual delivery meets the a priori definition -- a sort of index of integrity of treatment (Boruch & Gomez 1979; Yeaton & Sechrest, 1981). This index could serve as a variable in a regression equation, thereby providing an estimate of the influence of the treatment integrity on outcome criteria.

It is clear that by working with manuals, internal validity could be increased (at the cost of external validity), while stipulative definitions would permit conclusions that could be generalized to the practice of psychotherapy. Monitoring, which may be hard to accomplish, could increase the internal as well as the external validity of the study. The choice of the method for controlling the treatment will depend upon the goal of the evaluation. In critical research this control, a manipulation check, is best achieved through training manuals and monitoring. However, in an applied setting we agree with Cronbach (1982) who questions the sensibility for the applied scientist, e.g., the program evaluator, "to investigate a standard treatment that is unlikely to operate once the investigator's pressure for compliance is removed... Or should he permit, and collect data on, the heterogeneous interpretations
that schools place on the directive under normal operating conditions?" (p. 43).

But also clients have considerable degrees of freedom. Clients can boycott the treatment, more or less unnoticed by the therapist. Especially in research in which specific techniques are evaluated, the outcome data may be invalid if the clients do not practice the technique. Hoelscher, Lichstein, and Rosenthal (1984) investigated, by unobtrusively measuring the amount of time clients played their relaxation-tape, the extent to which clients complied with the relaxation practice. The data showed a great variability in individual (reported and actual) compliance. The authors stressed the point that researchers should assess the amount of actually received dose. Thus going further than Luborsky et al. (1982) who, in a pharmaco-therapeutic metaphor, stated that researchers should assess the amount of offered dose of the drug.

6. THE DESIGN AND ITS IMPLEMENTATION

Methodological rigor implies experimental control. In outcome research this is usually interpreted to mean the use of an appropriate control group and an adequate sample (APA-Commission, 1982; Meltzoff & Kornreich, 1970). Minimal treatment or waiting-list conditions are increasingly considered inappropriate, since they do not control for vital nonspecific and placebo influences. We think that this emphasis on control has resulted in groups of clients and therapists (non-clients and young therapists) such that these will, to a great extent, account for the fact that no large differences have been found between the different treatment conditions.

A recurrent critique is directed towards outcome research for not being representative of the practice of psychotherapy (Parloff, 1982). One may doubt the ecological validity, the extent to which the conditions of an experiment overlap with conditions in the field of psychotherapy (Brunswik, 1956). This is clearly illustrated by the following examples. First, Andrews and Harvey (1981) found that only a mere 22% of the studies analyzed by Smith et al. (1978), used clients who would normally have undergone psychotherapy. Secondly, even if clients who normally would have received psychotherapy are used, one must still doubt their representativeness. For instance, although the often cited study by Sloane et al. (1975) used only clients who were referred to an outpatient-clinic of a psychiatric hospital, they were mainly young, unmarried, and had a relatively high level of education.
Also the sample of clients in the Vanderbilt-study (Strupp & Hadley, 1979) was limited to male students between 17 and 24 years of age who, moreover, had received a letter to draw their attention to the possibility of receiving psychotherapy. In practice, clients are usually older, married, of lower socio-economic status (than at least Vanderbilt's students), and distressed to such an extent that they actually seek help. The same point holds for the therapists in outcome studies: Both Smith et al. (1980) and Shapiro and Shapiro (1982) reported an average of about 3 years of experience; the level of a postgraduate student. Although research has not shown any effects of the therapist's level of experience on outcome, these studies suffer from so many methodological problems (Parloff, Waskow, & Wolfe, 1978) that we cannot conclude that there really is no influence. Instead, we think that the relative inexperience of the therapists used in the studies contributes to the common results in outcome research, and the "inexperience" of the clients does too.

The lack of ecological validity is also apparent in other aspects of outcome research. O'Leary and Borkovec (1978), for instance, note that the 34 outcome studies reported in the Journal of Consulting and Clinical Psychology during 1976, employed an average of 5 sessions. This contrasts sharply with the findings of recent research (e.g., Lange, 1984) into the practice of psychotherapy in which an average length of 8 months of treatment for outpatients was reported, indicating an average number of sessions considerably greater than 5. Moreover, this length varied enormously (1 week - 48 months), which in itself is an interesting phenomenon.

Sloane et al. (1975) encountered another gap between research and practice. Originally, they thought that a period of 4 months of treatment would be sufficient. However, nearly 15% of their clients had some "concluding sessions" after this period. Moreover, at follow-up it appeared that 24 of their 60 clients had received additional therapy after these 4 months. As is now amply demonstrated (Lange, 1984), it is very hard to determine the exact moment at which therapy will in practice be terminated!

Another consequence of the homogeneous groups used in outcome research -- groups that are not representative of the variation that is present in the daily practice of psychotherapy -- is the fact that one can hardly get any information on indications for psychotherapy. We simply do not know which clients can be helped with the use of psychotherapy. Research aimed at clients who do not receive psychotherapy, or who drop out of treatment (attrition) could solve this problem. As
for attrition, this indeed is a big problem. Garfield (1980), for instance, reports drop-out percentages that vary between 30 to 65%, and Lange (1984) could not do any post-measurement on 40% of his cases. (It is informative to note that, in pre-treatment, these clients generally functioned worse socially than did those who remained in treatment.)

Considering the fact that studies have been using the same type of clients and therapists who, moreover, are not representative of psychotherapy in practice, it may be concluded that the negligible differences in effectiveness between differing therapeutic orientations is an artefact that should not come as a surprise.

7. THE ROLE OF PSYCHOTHERAPY OUTCOME RESEARCH

Although, with some academic distance, one would assume that the goal of a psychotherapy outcome study is to investigate the effectiveness of a particular treatment, each self-respecting school of psychotherapy has been trying to affirm (with almost missionary zeal) its superiority above competing schools by outcome studies -- the 'Grand-Prix'-issue (Gottman & Markman, 1978). While such zeal has served as a motivating force, it can be said that research on the outcome of psychotherapy received renewed impetus, not so much from methodological breakthroughs or feedback from practicing psychotherapists, but because of a development, initiated by those who fund psychotherapy, which has been called "the current climate of accountability" (Barlow, 1981; Garfield, 1981). If this assessment is correct -- and there is every reason to believe that it is -- we may expect a renewed interest in outcome research. Related to this is the development of meta-analysis. Whatever the difficulties associated with meta-analysis are, it allows us for the first time to derive impressively far-reaching conclusions regarding the effectiveness of different approaches to psychotherapy. And this is exactly what its connection with the issue of accountability is. Policy-makers will probably tend to pay more attention to a few simple conclusions, regardless of their validity, than to concentrate on the rather complicated and opaque original literature consisting of an enormous amount of individual studies, of which Smith et al. (1980) analyzed 375 and even they were criticized for not including a substantial amount! This problem has been succinctly stated by Parloff (1979) in the subtitle of his article: "A little knowledge may be a dangerous thing".
We notice an increasing awareness of the role as opposed to the goal of psychotherapy research, especially in a political context. As Scriven (1967) points out, the goal of an evaluation refers to the type of question people are trying to answer. In contrast, the role of evaluation refers to the use people want to make of the evaluation, to the reasons why it is done, and to its impact. Although role and goal are intertwined, the distinction may help to throw light on the problems that the psychotherapeutic community encounters in this era of accountability. The role of psychotherapy outcome research may undergo a great shift as we can already see in the development of meta-analysis. It is effectuated on studies, most of which were designed to investigate specific hypotheses about limited treatment conditions, whereas the meta-analysis aspires to draw general and often political conclusions, as is reflected in the title of the Smith et al. (1980) book: The benefits of psychotherapy.

We think that the classical paper by Edwards and Cronbach (1952) may shed some light on this matter. They distinguished between four types of research in psychotherapy: technique research, survey research, administrative or applied research, and critical research. As is summarized in Table 1, each of these has a different goal and plays a different role. Technique research plays an assisting role aimed at the development of instruments. Survey research can serve applied as well as theoretical purposes and is aimed at the exploration of relationships between variables. Administrative research should assist people in making decisions and is done in a political context. Critical research, finally, plays a role in theory development and testing. Its aim is to test precise hypotheses derived from theory.
Table 1. The roles and goals of research

<table>
<thead>
<tr>
<th>type of research</th>
<th>role</th>
<th>goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>technique</td>
<td>adjunctive</td>
<td>development of instruments</td>
</tr>
<tr>
<td>survey</td>
<td>applied</td>
<td>exploration of relationships</td>
</tr>
<tr>
<td></td>
<td>or theoretical</td>
<td></td>
</tr>
<tr>
<td>administrative</td>
<td>political</td>
<td>aid in decision-making</td>
</tr>
<tr>
<td>critical</td>
<td>theory</td>
<td>testing of hypotheses</td>
</tr>
<tr>
<td></td>
<td>development</td>
<td></td>
</tr>
</tbody>
</table>

We must conclude that psychotherapy outcome research is as diverse as the roles and the goals of research in general, each having its own place and its own function.

8. CONCLUSION

We may summarize the preceding pages in the following points.

(1) Instead of talking about the effect of psychotherapy, we should talk about the effects of psychotherapies, including improvement, deterioration, and negative effects. The different perspectives in evaluation should be reflected in the multiple outcome criteria, and we should not be surprised if they clash.

(2) We need a broader assessment of the context of the complaints, a distinction between therapeutic goals and target complaints, and an emphasis on the functioning of the client after therapy, without relying solely on the statistical significance of the differences between pre- and post-measurement.

(3) The more global and judgmental the data, the more often the magical number of a 2/3 improvement rate emerges (see Wojciechowski, 1984). We therefore need to specify the processes that affect the judgments.
(4) We should explain the method factor and not analyze it away. This is particularly important since it calls for a broader theoretical framework (and explanation) regarding "nonspecific" and placebo factors (see De Mey, this volume). Research in social psychology on demand characteristics and persuasion seems to offer the blueprints for the construction of such a framework.

(5) When comparing different treatments, a measure should be constructed to assess the integrity of the treatments. Otherwise it is unclear what is being compared.

(6) There is an enormous gap between outcome research and the practice of psychotherapy, particularly regarding such factors as characteristics of the clients and therapists, the timing of post-measurement, the attrition rate, and duration of therapy.

(7) Finally, we need to distinguish between the goal and role of outcome research. This distinction might result in at least two complementary strategies of research. One strategy should explicitly investigate the influence of the "not yet"-specific factors in psychotherapy on outcome and related issues, such as adherence or compliance. We refer to Rosenthal (1980) and Wilson (1980) for reviews of the type of studies we have in mind. The second strategy is "administrative" research, such as program evaluation (Broekman, 1984; Cronbach et al., 1980; Schippers et al., 1984): a more systematic description of psychotherapy as it is practiced; exploration of the need for (specific) treatment packages for target populations; research on the appropriateness and adequacy of the treatment; and, effort (cost) and efficiency (benefit) research.

REFERENCES


