1977

A Comparison of Self Ratings of Change Done Dependently and Independently

Pamela P. Irwin
Eastern Illinois University

PAPER CERTIFICATE #2

TO: 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.

6-1-77
Date

Author

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because_____

_____

_____

_____
Date

_____
Author

pdm

A Comparison of Self Ratings of Change

Done Dependently and Independently

(TITLE)

BY

Pamela P. Irwin

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Master of Arts

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

1977

YEAR

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

5/11/77
DATE

ADVISER

5/11/77
DATE

DEPARTMENT HEAD

A COMPARISON OF SELF RATINGS OF CHANGE DONE DEPENDENTLY AND INDEPENDENTLY

BY

PAMELA P. IRWIN

B. A. in Psych., Eastern Illinois University, 1975

ABSTRACT OF A THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts at the Graduate School
of Eastern Illinois University

CHARLESTON, ILLINOIS
1977

Abstract

A total of 243 Mental Hygiene and Introductory Psychology students completed ratings of adjustment in January of 1976. In May of the same year, 65 of the original Mental Hygiene students and 110 of the Introductory students completed the post-testing under one of four procedures. The post-testing procedures were: 1) dependent testing (defined as post-testing with awareness of pre-testing results) with an option to change their original test; 2) independent post-tesing which was the traditional method; then completion of dependent rating; 3) testing of their recall of original ratings, then dependent testing; and 4) dependent testing without the option to change the original. Change measures were then computed and analysed. The results showed: 1) no significant differences between Mental Hygiene and Introductory students; 2) significantly lower amounts of change in the dependent procedure than in the independent procedure; 3) significant differences between original testing and its recall with recall being higher than the original rating was; and 4) no significant difference among any of the dependent procedures. Results were discussed as to the implications toward future research in the area of evaluating psychotherapy and toward the development of new methods of measuring change. It was suggested that the most valid measure of change might be obtained by showing the subject his original rating and asking for his best possible estimate of change.

Acknowledgements

I would like to thank the members of my committee, Clayton Ladd (Phd.), Randall Best (Phd.), and William Kirk (Phd. pending) for their invaluable assistance throughout the completion of this study. I especially thank Clay Ladd for the friendship and warmth he put into his many hours of guidance. I would also like to thank Dr. John Rearden for all of the time and effort he gave in helping to compute and analyze the data. I am grateful for all of their advice and support.

I also would like to express my gratitude to the entire faculty of the Psychology department. Through them I have found many meaningful learning experiences and opportunities for growth.

A special thank you goes to my parents. Their support and encouragement in all areas of my life is deeply appreciated. I also wish to thank my in-laws for their frequent assistance in numerous areas which would have detained the completion of this study without their help.

And a very warm and loving thank you to my husband and son who have been in my heart every step of the way. Their presence in my life has often been the spark that kept me going.

Table of Contents

List of Tables

A Comparison of Self Ratings of Change Done Dependently and Independently

In the past fifteen years the field of psychology has expanded immensely. Whereas in 1961-62 there was a total of 12,191 students earning B.A., M.A., or Ph.D. degrees in psychological training programs, in 1973-74 there was a total of 60,080 students of psychology in the United States (Yearbook of Higher Education, 1975). Professionally, there are psychologists employed in areas ranging from the mental health institutions to various branches of the armed services with numerous other possible areas of employment (Rotman, 1968). With this wide expansion of the field, the necessity for meaningful research has also increased. If psychology is to be a valuable as well as necessary service as indicated by the growing interest in the field, then some measure of its utility must be continually examined.

In some areas of psychology the utility of the profession may be more observable than in other fields. If the industrial psychologist implements procedures whereby personnel incentive, and thus production, is increased, the value of his service is evident in observable dollar increased profits. If the school psychologist develops a method of teaching problem learners, their progress can be measured and examined. But what of the more ellusive areas of specialization?

As early as 1938, Landis (in Eysenck, 1952) attacked the problem of evaluating the effectiveness of psychotherapy noting problems with even determining valid outcome criteria with which to compare results. Eysenck (1952) was equally aware of the problems in evaluative psychotherapy research when he said, "There are obvious shortcomings in any actuarial comparison and these shortcomings are particularly serious when there is so little agreement among psychiatrists relating even the most fundamental concepts and definitions" (p. 323). But at the same time, the continued usefulness of psychotherapy itself would seem to be dependent upon the ability to measure its present usefulness. Malan (1973) goes so far as to say, "Outcome is surely the crucial variable in psychotherapy -- in the investigation of any therapeutic technique there is little point in studying other variables unless their relations to outcome can be established" (p. 719).

Therefore, the problem becomes one of determining what does psychotherapy actually do? If research is to complement and enhance clinical procedures then benefits or detriments of the various techniques must be discovered. Unfortunately, the almost limitless variables involved in clinical diagnosis, practice, and individual patient characteristics all serve to confound the task of specifying outcome results. The confusion in the early research on this topic (Malan, 1973) is ample evidence of the difficulties involved in this problem with more recent developments being equally important in discovering the present and future possibilities for meaningful research into the outcome problem.

Unfortunately however, the research to date on the

Problem of outcome evaluation has been fraught with many difficulties. Through examination of the various studies one finds repeatedly that controversy has arisen concerning both actual findings and their implications. Out of the numerous difficulties involved, though, little has been done in the way of changing the research procedures themselves. In other words, though frequent problems have been encountered in the traditional evaluation methodology (i.e. pre-test -- post-test to ascertain amount of change during the treatment time period) this same methodology has been utilized recurrently almost as if without question. Needless to say, the numerous difficulties involved in outcome evaluation cannot be relegated to deficiencies in the methodology alone. Yet, perhaps, examination of alternative methodological procedures might lead to new insights into the evaluation research in general. By examination from a new angle of the question the least that may be determined is that the traditional methodology is not at fault in the repeated occurrence of problems. However, without this examination having taken place, it does not seem particularly instrumental to continue past patterns without any question as to their possibly detrimental effects. Therefore, it is the purpose of this study to begin an examination of the utility of the traditional evaluation methodology while proposing a possible new procedure for examination of treatment effects.

Review of the Literature

The Psychotherapy Outcome Research

One of the most controversial issues in the area of outcome research began with Eysenck (1952) in his study, "The Effects of Psychotherapy: An Evaluation." In this study he compared available outcome research dealing with patients diagnosed as neurotic to come up with the following figures: "Patients treated by means of psychoanalysis improve to the extent of 44 per cent; patients treated eclectically improve to the extent of 64 per cent; patients treated only custodially or by general practitioners improve to the extent of 72 per cent. There thus appears to be an inverse correlation between recovery and psychotherapy; the more psychotherapy, the smaller the recovery rate" (Eysenck, 1952, p. 322). He felt that the results failed to support the hypothesis that psychotherapy fecilitates recovery from neurotic disorder but warned against over-generalization of these findings due to the numerous shortcomings throughout the available research in this field.

Landis (1937, in Rosenzweig, 1954, p. 298) had summarized some of these difficulties as -- "ignorance of the nature or cause of mental disease, disagreement among experts concerning even such broad differentiations as somatogenic or psychogenic, lack of uniformity with respect to categories of improvement -- and concludes: 'Because of these difficulties, it is apparent that statistical figures, rates of recovery, etc. have to be evaluated cautiously, precisely, and with a minimum of generalization.'" Though both Eysenck and Landis were well aware of the difficulties, Rosenzweig (1954) felt that Eysenck especially was

guilty of the over-generalization he had warned against.

Rosenzweig (1954) found many other problems in Eysenck's original position which he discussed as: 1) problems in comparing within a group diagnosed as neurotic when no specific definition of this classification exists; 2) problems in determining the extent and similarity of procedures of psychotherapy within both the control and experimental groups (i.e. does the control group control?); and 3) problems in defining improvement or the extent of recovery such that the criteria are comparable between groups. In each of these areas he found serious fault with Eysenck's work such that he concluded, "To undertake an evaluation of the effects of psychotherapy by tallying outcomes at second hand, without even introducing the problem of dynamic change in various forms of illness and in differing therapeutic procedures, and in default of such considerations, to reassign diagnoses and prognoses is to invite the inconsistencies and 'non-sequiturs' that have been demonstrated in the foregoing reanalysis" (Rosenzweig, 1954, p. 303).

The outcome issue continued with Baron and Leary's (1955) study, "Changes in Psychoneurotic Patients With and Without Psychotherapy." Using 150 psychoneurotic patients drawn from the same clinic population, they devided these into the experimental and the waiting-list control groups. Pre and post testing on the MMPI was used to determine what changes, if any, occurred in the profiles of patients who received psychotherapy as compared to the waiting-list controls who remained untreated during the same period of time. Their comparison with waiting-list controls differed from other such studies at that time (Kaufman,

1950; Gallagher, 1953; Mosak, 1950; and Rashkis and Shaskin, 1946) who had reported observations of MMPI changes following psychotherapy. Baron and Leary (1955) likewise found improvement on the MMPI of patients undergoing therapy but this was also seen in the waiting-list controls such that there was no significant difference between the two groups thus again questioning the effectiveness of psychotherapy.

Desmond S. Cartwright (1956) then attacked the Baron and Leary study on two important aspects. His examination of their material suggested that: 1) the samples differed as to the dispersion of changes (SDs) suggesting that the two samples were not drawn from the same parent population in regard to the variance of changes on some of the MMPI scales; and 2) that the greater amount of change was not all in one direction such that mean differences between the groups were absent because differences of two kinds, positive and negative, occurred. In fact, he determined that some of the therapy patients actually deteriorated to a greater extent than did the waiting-list controls while at the same time some therapy patients did improve significantly more than the controls. Thus, the combined mean difference scores of the subjects would have been confounded by this two-directional change. Thereby, the confusion in the outcome research problem was continued.

Earlier, Cartwright (1955) had also attacked the spontaneous remission position of Eysenck (1952), Landis (1937), and Denker (1946). In this study he found difficulties in the reliability of the recovery criteria that had been used and the past assumption that there is one and only one experimental

method of demonstrating the effectiveness of psychotherapy. In questioning the practice of using extreme conditions of presence and absence of psychotherapy to measure its usefulness, Cartwright felt that a more important variable was the amount of intervention as defined by depth and extent characteristics rather than the temporal quality. The basic logic he saw in the area was that "if it is a true statement that 'as a result of manipulating certain independent variables, psychotherapeutic intervention is differentially effective in producing recovery from neurosis,' then the simpler statement that 'psychotherapeutic intervention is effective in producing recovery from neurosis' is also true, whether the assumed spontaneous recovery factors exist or not" (Cartwright, 1955, p. 294).

Other studies did show several remedial effects of psychotherapy of several kinds and in many degrees: change from dissonance to consonance between self-perception and the perception of the self by others (Sullivan, 1947); disappearance of physiological symptoms (Saul, 1948); increased self esteem (Raimy, 1948); decreased defensiveness (Haigh, 1949); dissolution of hallucinatory and delusional states (Fromm-Reichman, 1950); reduction in physiological tension as measured by several objective, projective, and rating techniques (Rogers and Dymond, 1954) to mention a few. Therefore, it would seem that the very method of investigation and comparison is a major and influential determinant of the results.

Bergin (1963) also became involved in this area with his reinterpretation of psychotherapy outcome research, "The Effects of Psychotherapy: Negative Results Revisited." Discussing the

entirely negative results of the Cambridge-Somerville Youth Study of delinquency prevention (Powers and Witmer, 1951, in Bergin, 1963) he found two major difficulties; 1) that the treatment involved the counselor's being a warm, interested friend of the boys rather than what would normally be considered actual psychotherapy and 2) there was no way of positively determining that the control subjects used were, in fact, genuine controls. According to Bergin a fault in the literature is that we have little knowledge as to what actually goes on in control groups. Accordingly, Frank (1961) and Gurin, Veroff, and Feld (1960) found that people who are seriously upset will seek help from clergymen, physicians, friends, teachers, etc. at a significantly greater frequency than from mental health professionals. Therefore, if control subjects are in actuality obtaining such aid then their comparison to therapy patients to support a spontaneous remission argument becomes rather erroneous and misleading. What would be more important it seems, is to determine the ingredients that are causing changes in both groups.

Furthermore, specific therapist qualities such as empathetic understanding, congruence, and unconditional positive regard (Truax, 1963) and his being democratic, nonauthoritarian, noninterpretive, and personal in his approach (Whitehorn and Betz, in Betz, 1962) were found to influence the therapists' ability to promote positive rather than negative change in his clients. Obviously, these characteristics are not sole property of those in the mental health field which makes it possible that the same influential variables are operating upon both the control and experimental groups and thereby confusing the outcome research data.

Bergin (1966) then carried the issue farther in his digest of psychotherapy findings, "Some Implications of Psychotherapy Research for Therapeutic Practice." From this examination of the research data he arrived at the following conclusions:

> Conclusion 1:
> Psychotherapy may cause people to become better of worse adjusted than comparable people who do not receive such treatment. (This phenomena is what Bergin then labled "The Deterioration Effect").
>
> Conclusion 2:
> a) It has been frequently replicated . . . that control Ss who do not receive psychotherapy change positively as a group with the passage of time. This is the so-called "spontaneous remission" effect.
> b) Three studies indicate that many of these disturbed persons who receive no formal psychotherapy seek and obtain help from various professional and non-professional sources . . .
>
> Conclusion 4:
> To date, the only school of interview-oriented psychotherapy which has consistently yielded positive outcomes in research studies is the client centered approach.
>
> Conclusion 5:
> Even when various sources of slippage and inadequacy are accounted for, interviews still do not generally produce very dramatic changes in people. . . . many types of people simply are not helped at all by this procedure.
>
> Conclusion 6:
> Studies of learning have thus far been very fruitful in generating principles and methods for promoting personality change. The cases presented and research studies reported provide more positive evidence of the usefulness of these methods than is the case in any form of traditional interview or dynamic psychotherapy, including client-centered therapy.
>
> (Bergin, 1966, p. 235-239)

This article and Bergin's sweeping conclusions were soon met by criticism (Braucht, 1970; Schuldt and Truax, 1970) most of which centered around the conclusion proposing the presence of a deterioration effect. Braucht (1970, p. 243) focused his

arguments with Bergin in the areas of "problems of criterion validity, lack of experimental control, and inadequate experimental methodology and design." Re-emphasizing many of the difficulties Bergin himself had noted, Braucht (1970) also pointed out that more thorough examination of the material brings out further discrepancies. For instance, in the Fairweather (1960) study cited by Bergin, material that he did not bring out showed that "'maladaptive' changes occurred only in the case of long-term psychotics and that the variance effect was largely due to these chronic psychotic patients" (Braucht, 1970, p. 296). Furthermore, Braucht (1970, p. 296) pointed out that "movement in either direction on a scale may be an adaptive change for the individual patient" and from Fairweather (1960, p. 22, in Braucht, 1970, p. 297) he quoted, "it might be clinically more meaningful if a directional prediction of adaptive change were made for each patient prior to therapy. Thus, positive change for a patient would be determined by the direction in which clinicians agree that the patient should move rather than the scale itself." Braucht is very much against the wholesale application of social and experimental psychology derived research designs feeling that research in psychotherapy should involve individual predictions and more stringent controls of the various confounding patient-therapist variables. This belief would correspond with Schudlt and Truax's (1970) critique of the Bergin article in which they also attacked the generality of his conclusions. They had found the deterioration phenomena to be influenced by the specific outcome measure, certain groups of therapists, and certain groups of

patients. Thus, the major problems with Bergin's stand seem to be its generality, the statistical rather than individual orientation, and the lack of control of various influencial factors.

But Bergin (1970) continued with his position by replying to Braucht in an article re-emphasizing his original evidence. Stating that the individually tailored criterion that Braucht and Bergin himself (Bergin and Strupp, 1969; and Strupp and Bergin, 1969) had advocated was a very desirable quality in research, he still felt it to be inapplicable to the research that he had considered. Basically he re-emphasized the evidence supporting the deterioration effect while replying to Braucht's arguments in such a way so both sides seem to have equally justifiable points of support and areas of question. As Bergin stated, there is evidence in the research of the existence of a deterioration effect. Equally true is Braucht's questioning of the methods and resulting supportive material. Thus, though much has been written and considered, nothing could be stated as absolute, undeniable fact.

The controversy was then picked up by May (1971) in his critical review of the literature. Like many of his predesessors, he again questioned the conclusion that psychotherapy causes a mixture of improvement and deterioration. Reviewing separately the experimental studies which had supported the variance phenomena in psychotherapy outcome, he restated previous discrepancies and also found new elements of contention. For example, in some of the major studies, those by Powers and Witmer (1951), Rogers and Dymond (1954), Baron and Leary (1955), Mink and Isaksen (1959), Cartwright and Vogel (1960), Cartwright (1961),

Wheeler, et. al. (1950), and Meszaros (1958) which had all found a deterioration effect, May (1971) cited major difficulties with their comparisons between the experimental and control groups. According to him, the two groups were not equally matched or even randomly assigned which would create what he called "bounded variation" in that the controls being closer to normal had less freedom to move on the scale. Thus, what appeared to be greater variance in the experimental subjects could really be a methodologically obtained limit on the controls as to the amount of variance which was initially possible.

In the studies by Fairweather, et. al. (1960), Rogers, et. al. (1967), and Carkhuff and Truax (1965) May (1971) questioned the results on the grounds that drugs had not been controlled such that their influence could not be separated from the therapy influences. In other studies that May cited (Truax, 1963; Frank, 1967; Truax and Carkhuff, 1966; Kiesler, et. al., 1967; and Kiesler, et. al., 1967) strong evidence toward the need to uncover influencial patient-therapist variables was brought out in a way that showed their consideration to be a necessary inclusion in the outcome research.

As long as these basic design problems brought out by May and others before him are not solved, it would seem that the findings of psychotherapy research are questionable. When each successive study lacks the necessary comparison controls, or contains any of the other oft-mentioned problems, then little objective evidence is available.

With no conclusion at hand in this problem area, a historical review by Malan (1973), "The Outcome Problem in

Psychotherapy Research," again questioned the negative results and the problems within the studies themselves. After a thorough examination of the data, Malan felt that the questionable aspects of the "negative" studies decrease their impact and that there is renewed hope in recent studies based on dynamically oriented criteria for developing supportive and practical research. With a plea for research that can be used in clinical practice, he said, "What is needed above all is a determined, ruthless, and dispassionate search for the truth". . . (Malan, 1973, p. 728).

With the confusion of the past forty years that this review has shown, Malan's plea for more research does seem to be what will be needed before definitive results are obtained. So far all that has been "established" has then been shown to be questionable on one ground or another. Those involved in the practice of psychotherapy would obviously like to be able to substantiate their "feelings" of usefulness with objective, unbiased data. Yet, to date, such a task has not been completed without the immediate rejoinder by some other researcher questioning some aspect of the results. This review has brought out many of the frequently cited problems with outcome research which have ranged from inadequate control of comparison groups to lack of comparable recovery criteria. What has seldom, if ever, been questioned though is the measurement procedure itself.

Basically, the overall question in outcome research seems to be: how has this person changed? And, due to what causes? Combining the two problems of the measurement of change and the determination of the change producing element has probably been

one of the confusing factors in the research. Before getting into the causal factors it would seem useful to study the question of the measurement of change as a separate entity. Thus simplifying the question at hand may, at least, lead to a base for further developments. Still, the measurement of change itself is a formidable task as is also evident in the research.

Measurement Problems in the Research

With the realization of the problems involved in psychotherapy outcome research, interest in the entire area of measuring change grew. Zax and Klein (1960, p. 435) put the issue this way, "Evaluatory research in psychotherapy is a most complex activity but an extremely important one if we are to understand more about the nature of what can bring about personality change. The practical and theoretical problems involved in acquiring Ss, developing meaningful controls, and making measurements are enormous. Add to these the question of what one should measure, that is, what criterion should be used, and the complexity is increased many times over." But then their paper concentrated on the criterion problem again, with little information being given as to the solutions of the other factors they had deemed important.

Yet, as Cross (1964) pointed out, the failure to show psychotherapy benefits, or even to come up with adequate change measures, may well be due to methodological or measurement, not treatment, failures.

Bergin and Strupp (1970) noted a renewed appreciation of the internal, or experiential, processes that require development of new technology for studying such a private experience. At

the same time, they emphasized the failures of standard experimental designs and statistical procedures to bring about fruitful inquiry. As they said, "Among researchers as well as statisticians, there is a growing disaffection from traditional experimental designs and statistical procedures which are held inappropriate to the subject matter under study" (Bergin and Strupp, 1970, p. 25). Thus, the problem does not seem to have changed much in the past 15 years since Stevenson (1959) noted that there had been more publications stating that there are problems in evaluating psychotherapy than there had been experimental attempts to study its effects.

Therefore, the purpose of this study was to investigate one of the basic research assumptions, namely, that pre and post ratings of change should be done independently, i.e. without the rater being informed of pre-ratings when doing the post-ratings. This may not be optimal utilization of the rater. Though many studies have used subjects as judges (Snyder, 1953; Tucker, 1953; Fielder, 1949; Cowen and Combs, 1950; Mower, et. al., 1953; and Shapiro, 1961, 1969, 1970; to mention a few), none have given the subject all the available information to maximize the measurement of change. In each case the subject provides pre and post measures of adjustment (of whatever type the specific procedure called for) and then some difference measure between these two is determined, presumably reflecting the amount of change which had taken place. However, this study hypothesized that this traditional method may itself cause some of the many problems so far encountered in the research. Accordingly, the procedure proposed was to first test the subject prior to treat-

ment, then at the post-testing session give the subject full access to his original ratings, having him determine his "post-test" level of adjustment in light of the original rating and with full awareness that the most important consideration was the amount of change which had taken place.

To do this, it was first necessary to consider some of the problems involved. Use of a self-rating scale was decided upon as a valuable way for the subject to express both his original level of adjustment in several areas, and through post-testing to express the amount of change in adjustment. Rating scales have the advantages of being concise, being adaptable to variable numbers of categories for rating, allowing separate ratings for each item, being interesting and simple for the raters, being able to use fairly naive judges, and as Cronbach (1960) noted, they generally require only a fraction of the administration and assessment time of other evaluation devices.

On the other hand, rating scales have also been found to have some shortcomings, a major one being that of reliability (Miller, 1954). In an extensive article concerning the reliability, validity, and objectivity of rating scales, Wittenborn (1972) cited many of the problematic areas and gave possible solutions. To increase the reliability, the following suggestions may be drawn from his article: 1) use terms that are familiar to the rater and that do not have various meanings; 2) do not use terms that may be so general that several specific behaviors could be implied; 3) do not use terms that are so vague as to permit more than one possible rating for a given symptomatic feature; 4) do not construct scales to refer to value judgements (i.e. "good,"

"questionable," "unacceptable") rather than specific referents; 5) have the steps of the scale refer to explicit referents rather than such words as "mild," "moderate," or "severe;" 6) use scales in which the steps are transitive; 7) use rating procedures which specify the conditions of their use, i.e. the conditions of observation, the duration of the observation period; 8) make it clear whether the rating is to be based upon the behavioral referant per se or on some concept of which the behavioral referant is illustrative; 9) do not have scales require the rating of complex inference; and 10) specify the context for the rated behavior or experience.

Wittenborn's discussion of objectivity vs. subjectivity focuses on the clinician as the rater thereby not being particularly applicable to the situation here presented in which self-ratings were used. His discussion of validity brings up the important point that "the validity of a score should be perceived in terms of the degree to which it provides those particular distinctions that are required by their purposes" (Wittenborn, 1972, p. 85). (Wittenborn's discussion of reliability and validity was considered with respect to this particular study throughout the devising of the scales, instructions, and procedures.)

Another possible problem area in the present methodology was the use of the subject as the judge of his change. As Garfield, Prager, and Bergin (1971) put it, "On the one hand, the client as the person with the problem and as the recipient of the treatment should be in a favored position to evaluate the outcome of treatment. On the other hand, the client, depending

on a variety of factors may also be influenced in how he judges improvement, such as not displeasing the therapist, the 'hello-goodby effect' (Hathaway, 1948), suggestion, and the need to rationalize his investment of time and money" (Garfield, Prager, and Bergin, 1971, p. 307).

Yet, if psychotherapy is to continue as a meaningful service, it seems reasonable that the recipients of the service are the ones who must perceive the benefits regardless of what the technological (behavioral, physiological, unobtrusive, etc.) measures of their improvement might reveal. From a former patient the situation was put as, "Only by considering the total experience of the patient can the mental health professional determine whether or not he is providing helpful and useful services," and "Evaluation should determine whether or not a person is better able to handle his life situation after treatment, not whether his label changes" (Hoshall and Friedman, 1975, p. 8).

Likewise, Luborsky (1971) emphasized the patient's experience in saying, "When a patient reports that he has changed in certain areas, his statement (especially for non-psychotic patients) has considerable 'face validity'. Furthermore, his (and the therapist's) estimate of the worth of the change (i.e. the 'improvement') should be known. The fact that a patient changes from status A to status B may not be sufficient for a criterion measure -- the importance to the participants needs to be known. A measure of insight, for example, may show an increase, but we also should demonstrate how valuable it is to the patient. Sometimes the change need not be a big one to be a

big improvement; a very small change in a critical area can make a crucial difference to the patient" (Brenman, 1952, in Luborsky, 1971, p. 317).

And earlier, Rosenzweig (1954) had also emphasized the importance of the subjective reports of patient and therapist noting that though quantification may not be as possible it is also the process and dynamics of the patient's experience which must be considered. Likewise, Bergin and Strupp (1970, p. 22) pointed out that "there may be an important lesson from scores of studies that almost any set of procedures in the context of a benign human relationship, presented to, or viewed by, the patient as having therapeutic value will result in psychological or behavioral change describable as therapeutic."

Related to this subjective experience is a question that Mintz (1972, p. 11) pointed out as being so far unexplained -- that "Ss who begin therapy at relatively healthy levels on any particular measure simply cannot change much on the measure. But how are 'small changes' at relatively high levels experienced?" Thus, when the experiential factor in change is to be considered, as seems necessary from recent research, the subject as a judge and his self-reports have research advatages which may out-weigh their drawbacks.

Though objective quantification may be statistically imperative in some cases, the literature to date has not proved this method to be totally fruitful in the evaluation of psychotherapy or even in determining valuable measures of individual change. Therefore, not withstanding the subjectivity and possible unreliability of self-ratings, the traditional methods using

tests or therapist's ratings or other measures have been shown to contain equally serious flaws (Bergin and Strupp, 1970). Perhaps new methods will be found to have their own shortcomings but any new measurement technique offers, if nothing more, at least another approach which may yield new insights into the difficult but crucially important problem of accurately measuring the worth of psychological services.

Method

Subjects

Subjects were 243 college students at Eastern Illinois University. No selection was made as to age or sex. The total number was comprised of 109 students enrolled in the Mental Hygiene classes and 134 students enrolled in the Introductory Psychology courses.

Apparatus

Seventeen rating components were drawn from 42 Ratings of Adjustment as devised by Ladd (1967). The particular scales were chosen to cover a wide variety of adjustment factors and to obtain items most likely to change within the 4 month time period of the study (i.e. Comfort -- from relaxed to anxious -- might well change within 4 months whereas a rating of Physical Health -- from excellent to poor -- would be less likely to show measureable change within the time limits). Each of the seventeen scales were described according to the two extreme levels by behavioral examples as were other considerations made to increase the reliability as suggested by Wittenborn (1972).

On the answer sheet was the 0-100 point rating scale to be used for all ratings. This scale with its unit normal curve specification of college students gave easy reference for the subject's comparisons. The 0-100 point scale was chosen to give the subjects an optimal measure on which to indicate change in the post-testing. With the large range of possible ratings both

large and small changes could be easily expressed.

An instruction sheet was also given to each subject; thus, the testing materials consisted of the instruction sheet, the scale description booklet, and the answer sheet. (See Appendix A for all materials used).

Procedure

All subjects were originally tested in their respective classes in January, 1976, to determine the initial levels of adjustment.

To test the hypothesis that the traditional "independent" (i.e. without awareness of original test scores) testing procedure may be a relatively inaccurate measure of change, the "dependent" test procedure was devised and defined as post-testing with awareness of original test status.

Prior to post-testing, subjects were randomly divided within classes into four groups such that comparative numbers of Mental Hygiene or Introductory students were represented in each of the four procedural groupings. The procedures for post-testing in these groups were as follows:

Group 1:

The post-testing was done following the dependent procedure (with awareness of original ratings) with the option for subjects to change their original test.

Group 2:

For comparison purposes this group completed two procedures:

A) Traditional independent post-testing; and

B) Dependent post-testing with the answer sheet of A

being removed prior to the B procedure.

Group 3:

To test the assumption that traditionally measured changes include a subjective memory effect (i.e. that the subject tries, whether overtly or covertly, to remember his original status when being given a repetition of the pre-test for the post-test, as is the traditional procedure, before he scores himself; and that faults in his memory will have an effect on the face validity of the change measure obtained) this group also completed two separate procedures:

A) Completed ratings according to instructions to re-rate their original (i.e. January pre-testing) level of adjustment to test the memory effect; and

B) Then removal of the A materials and completion of the dependent testing procedure.

Group 4:

The post-testing was done following the dependent procedure this time without the option to change the original pre-test.

The first three groups when going through the dependent procedure were given the option of changing their original score sheet so that the most accurate measurement of change could be obtained. It was felt that the case could arise in which a subject would be unable to give an accurate measure of the amount of change because he had incorrectly judged his original level of adjustment. For example, he might have originally scored his "Comfort" level at 90 but after working on improvements during the semester realized he was far more nervous at the original testing than he had recognized. In this way he could lower his original score and then

and then make a post-rating which would more accurately express the amount of change that he had experienced. This option of changing the original could not, of course, be given in the Independent (2-A) and Memory (3-A) procedures.

After completing all ratings, subjects in all groups were asked to circle on the answer form the numbers of two areas they had actively worked on during the semester. In this way comparisons of the amount of change on these items could be made with items of less significance to the subject.

Post-testing was also done in the respective classrooms. The only added verbal instructions were clarification of the written instructions if questions were asked, a reminder to circle the two items worked on, and instructions to put year in school and number of Psychology classes previously taken on their answer sheet in case this data was later needed for comparison purposes.

The following chart can be used to see the various procedures among the groups:

Testing Procedures

Group 1	Group 2	Group 3	Group 4
January	January	January	January
Pre-test	Pretest	Pre-test	Pre-test
May	May	May	May
Dependent post-test with option to change original.	Independent tradition-al post-test.	Memory-effect post-test.	Dependent post-test without option to change original.
	(Removal of materials from above procedures)		
Circle 2 items worked on.	Dependent post-test (with option to change original)	Dependent post-test	Circle 2 items worked on.
	Circle 2 items worked on.	Circle 2 items worked on.	

The total number in each of the groups (1 through 4) was comprised of one-half Mental Hygiene and one-half Introductory Psychology students for the January pre-test. The numbers of students in the May post-tests varied due to the fact that students

had withdrawn from the classes or were absent when the post-testing was done. The randomizations had been done prior to the final testing such that the groups were originally divided equally.

Hypotheses and Statistics

1) The first comparison was between the Mental Hygiene students and the Introductory Psychology students. It was hypothesized that no significant difference would be found between the overall amount of change in these groups. An analyses of variance was used to compare the overall changes.

2) It was hypothesized that the circled 2 items worked on would show a significant difference in the amount of change between the Mental Hygiene and Introductory classes. A T-test was used to compare the differences.

3) In the Group 2 procedure it was hypothesized that when given the opportunity to know the original rating and to make the dependent rating, the change score will be different from that obtained in the independent procedure. It was clinically expected that the dependent procedure would show smaller changes due to the probability of obtaining exagerated change scores in the independent procedure when the subject is guessing as to his original level. The T-test for correlated samples was used to determine if the dependently and independently determined change scores were significantly different.

4) In the Group 3 procedure it was hypothesized that the original level ratings would differ from the memory-effect levels. A T-test was used to compare the D score differences between these procedures. The mean ratings were also determined so that the differences could be examined.

5) It was hypothesized that the overall amount of change obtained under the dependent procedures in all experimental conditions would not be significantly different. Analysis

of variance was used to compare these scores.

6) It was hypothesized that the option to change the original would have no significant effect upon the amount of change obtained throughout the procedures. An analysis of variance was used to compare the means of these groups.

Results

A total of 175 of the original 243 persons completed both pre and post testing. Of these 175 people, 110 were in the Introductory Psychology classes and 65 were in the Mental Hygiene classes.

The focus of this study was on measuring change over time. The amount of change for each individual was determined by using the D score (Cronbach, 1953, 1958; Cook, 1966; Mendelsohn, 1967). Thus, each subject's responses to the 17 ratings were converted to a single overall index of change, regardless of direction. There were either one or three D scores for each subject depending upon the procedural grouping he was in (i.e. Groups 1 and 4 each had one D score per person while Groups 2 and 3 each had three D scores per person due to the extra treatments in their groupings). These D scores were used for all statistical evaluations. See Appendix B for an example of the method of calculating D scores.

It was hypothesized that no significant difference in amount of overall change would be found between the Introductory and Mental Hygiene classes. Past research has indicated that when an overall change measure is used it is less likely to find marked differences between groups. Thus, it was necessary in this study to determine if any differences existed between the classes as to amount of overall change obtained.

To test the hypothesis that no significant difference in overall amount of change would be found, an analysis of variance was used. The first analysis included the D scores from the Independent procedure in Group 2. As can be seen in the following table, no significant difference was found in amount of overall change between these groups:

Table 1

An Analysis of Variance of Amount of Overall Change in Mental Hygiene and Introductory Psychology Students (Inclusion of Independent D scores for Procedure 2)

Source of Variation	DF	Mean Sign.	F	Sign. of F
Main effects	4	697.941	.785	.536
Course	1	0.10	.000	.997
Group	3	930.428	1.047	.373
2-way interaction				
Course-Group	3	1574.277	1.771	.155
Residual	167	888.785		

Then to double check these results another analysis of variance was computed this time including the dependent D scores from Group 2. In this way, all procedures were covered in the analyses to determine the effects of being in either of the classes. Again, no significant difference was found as can be seen in the following table:

Table 2

An Analysis of Variance of Amount of Overall Change in Mental Hygiene and Introductory Psychology Students (Inclusion of Dependent D scores for Procedure 2)

Source of Variation	DF	Mean Sign.	F	Sign. of F
Main Effects	4	114.916	.126	.973
Course	1	.907	.001	.975
Group	3	151.480	.167	.919
2-way interaction				
Course-Group	3	1575.196	1.733	.162
Residual	167	908.719		

Thus, it can be seen that there was no significant difference between the Mental Hygiene and Introductory Psychology students in their overall amount of change.

The second hypothesis was that Mental Hygiene students gained more on items defined as ones they had actively worked on than would Introductory students. This is due to the fact that the Mental Hygiene class itself was concerned with and encouraging towards psychological self-help and improvement whereas the Introductory class was not. Unfortunately, only a total of 25 students (Introductory N = 12; Mental Hygiene N = 13) circled any items at all with many of them circling only one item. Thus, a T-test was not used but the mean amount of gain on these items per class grouping was determined. The mean

gain for the Introductory students was 25.95 points while the mean gain for the Mental Hygiene students was 16.88.

Thirdly, it was hypothesized that the amount of change in the dependent rating procedure (the rater knows how he rated himself before) in Group 2 would be significantly smaller than that in the independent rating procedure (rater does not know previous rating). It was expected that the lack of awareness of one's original scores in the independent procedure would encourage students to escalate their post-test scores in an attempt to show they had changed. A T-test was used to compare these two procedures for measuring change. A highly significant (.001) difference of 5.11205 was obtained. The dependent procedure produced smaller differences than in the independent procedure, as can be seen in the following table:

Table 3

T-test for Comparison of

Independent and Dependent Procedures

Group	Number	Mean	Standard Deviation
Independent	38	66.1579	26.412
Dependent	38	54.4211	28.1537
Mean difference		T-ratio	
11.7368		5.11205 on 37 df	

The fourth hypothesis was that the actual pre-test ratings would differ from the post-test recall of the original ratings. It was thought that people would be unable to remember their pre-test ratings. If so, this would suggest that change measures obtained in dependent (e.g. "How much have you

changed?") post-testing would not show the same perceptions of change as would be obtained in traditional pre-post testing. A T-test was used to compare the D scores between the original-memory differences and the original-dependent differences. It was found that there was a significant difference at the .01 level of 2.86395 as can be seen in the following table:

Table 4

T-test for Comparison of

Original-Memory and Original-Dependent Procedures

Group	Number	Mean	Standard Deviation
Orig.-Mem.	41	55.8293	26.3058
Orig.-Dep.	41	43.122	23.3818
	Mean difference	T-ratio	
	12.7073	2.86395 on 40 df	

Next, the mean original rating of each individual and his mean recall rating were computed. Then an overall mean for all subjects in the original procedure and one for all of their recall ratings was determined. In this way the direction of the faulty memory could be determined. The overall comparison showed a mean of 74.9 in the original ratings and a mean of 80.07 in the recall ratings. Also, it can be noted that out of the total 41 people only 8 subjects had lower mean recall than mean original ratings, 2 had equal mean ratings (though item by item memory was still faulty), and the remaining 31 individuals perceived themselves in the post-testing as having been at higher original levels. For examination of individual mean ratings, see

Appendix C.

Fifth, it was hypothesized that the overall amount of change under the dependent procedures would not differ. To examine this hypothesis the needed analysis was of whether or not having seen the original ratings had any significant effect. Thus, a one-way analysis of variance was run contrasting Procedure 2-A with the other 3 procedures. The following calculations were derived:

Table 5

One-way Analysis of Variance

	St. error	T value	DF	T Prob.
Pooled Variance Est.	16.500	-1.757	171	.081
Separate Variance Est.	15.052	-1.926	170	.056

Thus, it was determined that the fact of seeing the original results did not make any of the groups significantly differ from the others such that the overall amount of change did not differ among the groups.

Sixth, it was hypothesized that the option to change the original pre-test would have no significant effect upon the amount of change. To determine this a one-way analysis of variance was run contrasting the first 3 groups (who had the option to change) with the 4th group (who was not given the option to change). The following calculations were derived:

Table 6

One-way Analysis of Variance

	St. error	T value	DF	T Prob.
Pooled Variance Est.	15.537	-.330	171.0	.742
Separate Variance Est.	16.383	-.313	168.9	.755

Thus, it can be seen that having the option to change the original test did not significantly effect the amount of change obtained.

Due to the expectations of accepting the null hypothesis in many of these cases, significance was checked at even the most liberal of levels and was not found at these levels.

Descriptive statistics were also calculated for each of the various sub-groups. The divisions were by class and procedures such that there were a total of 8 groups (G1-1; G1-2; G1-3; G1-4; G2-1; G2-2; G2-3; G2-4) with the first number representing the class (1 = Introductory; 2 = Mental Hygiene) and the second number representing the procedural grouping. For examination of these statistics see Appendix D.

Discussion

It was not originally expected that any significant difference would be found between the Mental Hygiene and Introductory students even though the Mental Hygiene students were involved in a classroom experience more directed toward influencing self-change. Past research has repeatedly indicated little change when the measure is an overall quantity on a number of variables as was true in this study. Also, with the relatively short time period (about 4 months) between pre and post testing, a large overall change in one group over the other would probably be unlikely. Again, this hypothesis and its results may be a further indication of the problems in attempting to define and examine change with an overall measure rather than specific item by item measures. It may be that overall measures are necessary for statistical comparisons. However, in the practical therapeutic atmosphere, the implication would seem to be that individualized considerations of change by item or by subjective importance rather than by one overall measure would also be useful.

Accordingly, it was thought that in Hypothesis 2 the gains per items worked on would show more improvement in Mental Hygiene students who were supposedly more involved in experiences influencial toward change than in Introductory students. Again, the scanty response to the directions to circle two items worked on was unfortunate. But still, the large difference between means and especially the direction of this difference is interesting

and indicative of a need for further study. Perhaps the more restrained estimates of change in the Mental Hygiene students is further evidence that heightened awareness may lead to less expansive perceptions of amount of change. Another possibility is that Introductory students are Freshman and making many adjustments to a new environment. However, any explanation with so little data is highly speculative; but the above results do show an interesting avenue for further experimentation and examination.

Hypothesis 3 and its implications represent the core of this study. Traditionally it has been assumed that for objectivity the pre and post testing would have to be done independently. The basis of this assumption was that if shown the original tests and asked to determine the amount of change through the post-test ratings, the subjects would necessarily inflate the amount of change. In fact, this experimenter was repeatedly warned that such "subjective" responding would be inflated and thereby invalid. Also, it seems that much of the research to date has operated under the assumption of the necessity of objectivity to avoid exagerated change measures even though continued problems arose with these methods. Thus, it is just this basic assumption which comes into question when it was found that knowledge of the original ratings led to significantly smaller perceptions of change.

One explanation of these findings could be as follows: When a person is given pre and post testing he automatically assumes that the tester is attempting to measure his amount of change in the elapsed time period. In the independent measure

he is unaware of his exact original level but may have in mind that he changed since the original testing. Thus, being aware of his faulty recall of the initial level he exagerates the amount of rated change to be sure to reflect the actual change he feels he has made. Thus, the faulty memory produces an exagerated estimate of change.

Likewise, the traditional pre and post testing method asks for ratings at the particular moment of testing and then assumes that the difference between these amounts is equal to the amount that the person has in fact changed. Though usually considered as being objective, both ratings are still subjective. Therefore, if subjective ratings are seen as valuable, then cannot the subjective estimate of change also be seen to be a valuable tool of study? In fact, the data here suggests that having more trust in the rater's subjective judgement by disclosing original ratings increases the validity of the outcome measure.

Without significant differences between Mental Hygiene students and Introductory students, the significant differences of the two testing procedures would seem to indicate that the methodological testing procedure does in fact influence the final outcome of any particular study. Thus, the utility of further examination of procedural effects can be seen as a useful approach to the outcome research which past data has frequently failed to include. If it is repeatedly determined that traditional methods produce significantly higher change measures than do dependent testing it may be grounds for questioning the past avoidance of subjectivity. At the very least, it seems

reasonable to utilize the rater himself to the utmost of his potential by giving the increased awareness as in dependent testing and then to compare these results with other outcome criteria. Though beyond the scope of this study, such new approaches to the question of outcome research may lead to valuable insights to a controversial yet equally important field of research.

As brought out in the discussion of Hypothesis 3, it was proposed that in traditional pre and post test procedures, subjects may attempt to remember pre-test levels when making the post-ratings. Then if memory of initial testing levels is incorrect, the estimate of change will be effected to the extent of the error of the memory. Assuming that memory does play at least some part in the post test levels, it seems reasonable to examine one's ability to remember these initial levels. The results of this study suggest that memory is quite influencial.

In examination of the direction of the faulty memory as herein found, it can be seen that most subjects perceived themselves as having been at higher original levels. Therefore, in traditional pre-post testing the escalated recalled adjustment would lead to the exagerated change measure as was found here. Likewise, if the recalled adjustment is an underestimate of the original rating, then the independently arrived change score will be an under-estimate of the person's improvement.

It may be questionable as to whether or not memory does play a part in the subject's post-test responding, but if it is at all possible that it does, then the above results would indicate a need for control of this influencial factor in future

research. If memory deficiencies are related to exagerated change measures as found in the independent procedure, a possible support for the use of dependent testing may be found and thereby add a new approach to the area of outcome research.

The fifth and sixth hypotheses and examinations were necessary to exclude extraneous influences of having various new procedures. The results of the fifth hypothesis showed that the dependent groups were consistant such that there was not an effect from having seen the original test.

The procedure of allowing subjects to change their original test was to further utilize the rater by giving him as much opportunity as possible in determining his best, most accurate estimate of change. The results indicated that having this option was not influencial in effecting the amount of change. Therefore, since so few subjects (20%) changed these scores when given the opportunity, the optimal procedure might be to simply show the subject his original ratings and request his best possible estimate of change.

Though the present study appears to be a somewhat radical departure from traditional methodologies, it is proposed with the utmost reserve. Years of developments in outcome research have led to many useful practices. Yet, some procedures have become standard practice without evidence that they are the best methods. It is the contention of this researcher that new developments and further studies may lead to new methodologies. These results would seem to imply the need for further experimentation whereby the ultimate usefulness of such new procedures may be determined.

References

Baron, R. and Leary, T. F. Changes in psychoneurotic patients with and without psychotherapy. Journal of Consulting Psychology, 1955, 19 (4), 239-245.

Bergin, Allen E. The deterioration effect: A reply to Braucht. Journal of Abnormal Psychology, 1970, 75 (3), 300-302.

Bergin, Allen E. The effects of psychotherapy: Negative results revisited. Journal of Counseling Psychology, 1963, 10, 244-250.

Bergin, Allen E. Some implications of psychotherapy research for therapeutic practice. Journal of Abnormal Psychology, 1966, 71 (4), 235-246.

Bergin, Allen E. and Strupp, Hans H. New directions in psychotherapy research. Journal of Abnormal Psychology, 1970, 76 (1), 13-26.

Braucht, G. Nicholas. The deterioration effect: A reply to Bergin. Journal of Abnormal Psychology, 1970, 75 (3), 293-299.

Brenman, M. On teasing and being teased and the problem of "moral masochism." In Luborsky, Lester. Perrenial mystery of poor agreement among criteria for psychotherapy outcome. Journal of Consulting and Clinical Psychology, 1971, 37 (3), 316-319.

Carkhuff, R. R. and Truax, C. B. Lay mental health counseling: The effects of lay group counseling. Journal of Consulting Psychology, 1965, 29, 426-431.

Cartwright, D. S. Effectiveness of psychotherapy: A critique of the spontaneous remission argument. Journal of Consulting Psychology, 1955, 2 (4), 290-296.

Cartwright, D. S. Note on "Changes in psychoneurotic patients with and without psychotherapy." Journal of Consulting Psychology, 1956, 20 (5), 403-404.

Cartwright, D. S. and Vogel, J. L. A comparison of changes in psychoneurotic patients during matched periods of therapy and no therapy. Journal of Consulting Psychology, 1960, 24, 121-127.

Cartwright, R. D. The effects of psychotherapy on self-consistency: A replication and extension. Journal of Consulting Psychology, 1961, 25, 376-382.

Cook, T. E. The influence of client-counselor value similarity on change in meaning during brief counseling. Journal of Counseling Psychology, 1966, 13, 77-81.

Cowen, E. L. and Combs, A. W. Follow-up study of 32 cases treated by non-directive psychotherapy. Journal of Abnormal Social Psychology, 1950, 45, 232-258.

Cronbach, L. J. Essentials of Psychological Testing. New York: Harper and Rowe, 1960.

Cronbach, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri and L. Petrullo (Eds.) Person, Perception and Interpersonal Behaviors. Stanford: Stanford University Press, 1958, 353-380.

Cronbach, L. J., and Gleser, G. C. Assessing similarity between profiles. Psychological Bulletin, 1953, 50, 456-473.

Cross, H. J. The outcome of psychotherapy: A selected analysis of research findings. Journal of Consulting Psychology, 1964, 28, 413-417.

Denker, P. G. Results of treatment of psychoneurosis by the general practitioner. A follow-up study of 500 cases. New York State Journal of Medicine, 1946, 46, 2164-2166.

Eysenck, H. J. The effects of psychotherapy: An evaluation. Journal of Consulting Psychology, 1952, 16, 319-324.

Fairweather, G. W., Simon, R., Gebhard, M. E., Weingarten, E., Holland, J. L., Sanders, R., Stone, G. B., and Reahl, J. E. Relative effectiveness of psychotherapeutic programs. Psychological Monographs, 1960, 74 (5, Whole No. 492).

Fielder, F. E. An experimental approach to preventive psychotherapy. Journal of Abnormal Social Psychology, 1949, 44, 386-393.

Frank, J. D. Does psychotherapy work? International Journal of Psychiatry, 1967, 3, 153-155.

Frank, J. D. Persuasion and Healing. Baltimore: Johns Hopkins Press, 1961.

Fromm-Reichman, Frieda. Principles of Intensive Psychotherapy. Chicago: University of Chicago Press, 1950.

Gallagher, J. J. MMPI changes and client-centered therapy. Journal of Consulting Psychology, 1953, 17, 334-338.

Garfield, Sol L., Prager, Richard A., and Bergin, Allen E. Evaluation of outcome in psychotherapy. Journal of Consulting and Clinical Psychology, 1971, 37 (3), 307-313.

Gurin, G., Veroff, J., and Field, S. Americans View Their Mental Health. New York: Basic Books, 1960.

Haigh, G. Defensive behavior in client-centered therapy. Journal of Consulting Psychology, 1949, 13, 181-189.

Hathaway, S. R. Some considerations relative to nondirective psychotherapy as counseling. Journal of Clinical Psychology, 1948, 4, 226-231.

Hoshall, Don and Friedman, Joel. Evaluation from a former patient's point of view. Evaluation, 1975, 2 (2), 8-9.

Kaufman, P. Changes in the Minnesota Multiphasic Personality Inventory as a function of psychiatric treatment. Journal of Consulting Psychology, 1950, 14, 458-464.

Kiesler, D. J., Klein, M. H., Mathieu, P. L. and Schoeninger, D. Constructive personality change for therapy and control patients. In Rogers, C. R., Gendlin, E. T., Kiesler, D. J. and Truax, C. B., (Eds.) The Therapeutic Relationship and Its Impact. Milwaukee: University of Wisconsin Press, 1967, 251-294.

Kiesler, D. J., Mathieu, P. L., and Klein, M. H. A summary of the issues and conclusions. In Rogers, C. R., Gendlin, E. T., Kiesler, D. J., and Truax, C. B., (Eds.) The Therapeutic Relationship and Its Impact. Milwaukee: University of Wisconsin Press, 1967, 295-311.

.add, C. E. 1967, Unpublished material.

.andis, C. Statistical evaluation of psychotherapeutic methods. 1937. In Eysenck, H. J. The effects of psychotherapy: An evaluation. Journal of Consulting Psychology, 1952, 16, 319-324.

.uborsky, Lester. Perrenial mystery of poor agreement among criteria for psychotherapy outcome. Journal of Consulting and Clinical Psychology, 1971, 37 (3), 316-319.

Malan, D. H. The outcome problem in psychotherapy research. Archives of General Psychiatry, 1973, 29, 719-729.

May, Phillip R. A. For better or for worse? Psychotherapy and variance change: A critical review of the literature. Journal of Nervous and Mental Disorders, 1971, 152, 184-192.

Mendelsohn, G. A., and Geller, M. H. Similarity of missed sessions and early terminations. Journal of Counseling Psychology, 1967, 14, 210-215.

Meszaros, A. F. and Gallagher, D. L. Measuring indirect effects of treatment on chronic wards. Disease of the Nervous System, 1958, 19, 167-172.

Miller, J. G. Criteria and measurement of change during psychiatric treatment. Bulletin of the Menninger Clinic, 1954, 18, 130-137.

Mink, O. G. and Isaksen, H. L. A comparison of effectiveness of nondirective therapy and clinical counseling in the junior high school. School Counselor, 1959, 6, 12-14.

Mintz, Jim. What is "success" in psychotherapy? Journal of Abnormal Psychology, 1970, 76 (1), 13-26.

Mosak, H. H. Evaluation of psychotherapy: A study of some current measures. Unpublished doctor's dissertation, University of Chicago, 1950.

Mower, O. H., Light, B. H., Luria, Zella, and Zeleny, Marjorie P. Tension changes during psychotherapy with special reference to resistance. In O. H. Mower (Ed.), Psychotherapy: Theory and Research. New York: Ronald Press, 1953, 546-640.

Powers, E. and Witmer, Helen. An experiment in the prevention of delinquency. 1951. In Bergin, Allen E. The effects of psychotherapy: Negative results revisited. Journal of Counseling Psychology, 1963, 10, 244-250.

Raimy, U. C. Self-reference in counseling interviews. Journal of Consulting Psychology, 1948, 12, 153-163.

Rashkis, H. A. and Shaskin, D. A. The effects of group therapy on personality scores. American Journal of Orthopsychiatry, 1946, 16, 343-349.

Rogers, C. R. and Dymond, Rosalind F. Psychotherapy and Personality Change. Chicago: University of Chicago Press, 1954.

Rogers, C. R., Gendlin, E. T., Kiesler, D. J. and Truax, C. B. (Eds.) The Therapeutic Relationship and Its Impact. A Study of Psychotherapy in Schizophrenics. Milwaukee: University of Wisconsin Press, 1967.

Rosenzweig, Saul. A transvaluation of psychotherapy: A reply to Hans Eysenck. Journal of Abnormal Social Psychology, 1954, 49, 298-304.

tman, Bertram T. Careers in Psychology. Washington, D. C.: B'Nai B'Rith Vocational Service, 1968.

ul, L. J. The physiological effects of psychoanalytic therapy. In F. Alexander and T. M. French (Eds.). Studies in Psychosomatic Medicine. New York: Ronald Press, 1948, 91-100.

huldt, W. John, and Truax, Charles B. Variability of outcome in psychotherapeutic research. Journal of Counseling Psychotherapy, 1970, 17 (5), 405-408.

apiro, M. B. A method of measuring psychological changes specific to the individual psychiatric patient. British Journal of Medical Psychology, 1961, 34, 151-155.

apiro, M. B. Short-term improvements in the symptoms of affective disorder. British Journal of Social and Clinical Psychology, 1969, 8, 187-188.

apiro, M. B. et. al. The Reduction of Mental Distress. London: Tavistock, 1970.

yder, W. Y. Group Report of a Program of Research in Psychotherapy. State College, Pa: Pennsylvania State College, 1953.

evenson, I. The challenge of results in psychotherapy. American Journal of Psychiatry, 1959, 116, 120-123.

rupp, H. H. and Bergin, A. E. Some empirical and conceptual bases for coordinated research in psychotherapy: A critical review of issues, trends, and evidence. International Journal of Psychiatry, 1969, 7, 18-90.

Sullivan, H. S. Conceptions of Modern Psychiatry. Washington, D. C.: William Alanson White Psychiatric Foundation, 1947.

Truax, C. B. Effective ingredients in psychotherapy: An approach to unraveling the patient-therapist interaction. Journal of Counseling Psychology, 1963, 10, 244-250.

Truax, C. B. and Carkhuff, R. R. Toward Effective Counseling and Psychotherapy: Training and Practice. Chicago: Aldine, 1966.

Tucker, J. E. Measuring client progress in client-centered therapy. In W. U. Snyder (Chmn.), Group Report of a Program of Research in Psychotherapy. State College, Pa.: Pennsylvania State College, 1953.

Wheeler, E. O., White, P. D., Reed, E. W. and Cohen, M. E. Neurocirculatory asthenia (anxiety neurosis, effort ayndrome, neurasthenia): A twenty year follow-up study of 173 patients. Journal of the American Medical Association, 1950, 142, 877-889.

Whitehorn and Betz, Barbara J. in Betz, Barbara J. Experience in research in psychotherapy with schizophrenic patients. In H. H. Strupp and L. Luborsky (Eds.), Research in Psychotherapy. Vol. 2, Washington, D. C.: American Psychological Association, 1962, 41-60.

Wittenborn, J. R. Reliability, validity, and objectivity of symptom-rating scales. The Journal of Nervous and Mental Disease, 1972, 154 (2), 79-87.

Yearbook of Higher Education. Chicago: Marquis Academic Media, 1975.

Zax, Melvin and Klein, Armin. Measurement of personality and behavior changes following psychotherapy. Psychological Bulletin, 1960, 57 (5), 435-448.

Appendix A -- Parts A_1 - A_6

Testing Materials

Part A_1 -- Answer form - used in all ratings.

Part A_2 -- Instruction sheet - used in original pre-test and procedure 2-A in post-test.

Part A_3 -- Instruction sheet - Used in post-testing procedures 1, 2-B, and 3-B.

Part A_4 -- Instruction sheet - used in post-testing procedure 3-A.

Part A_5 -- Instruction sheet - used in post-testing procedure 4.

Part A_6 -- Rating scale description booklet - used in all ratings.

RATINGS OF ADJUSTMENT -- ANSWER FORM

Name:_____

Date:_____

Age:_____

Sex:_____

Course:_____

Scale 1: _____

Scale 2: _____

Scale 3: _____

Scale 4: _____

Scale 5: _____

Scale 6: _____

Scale 7: _____

Scale 8: _____

Scale 9: _____

Scale 10: _____

Scale 11: _____

Scale 12: _____

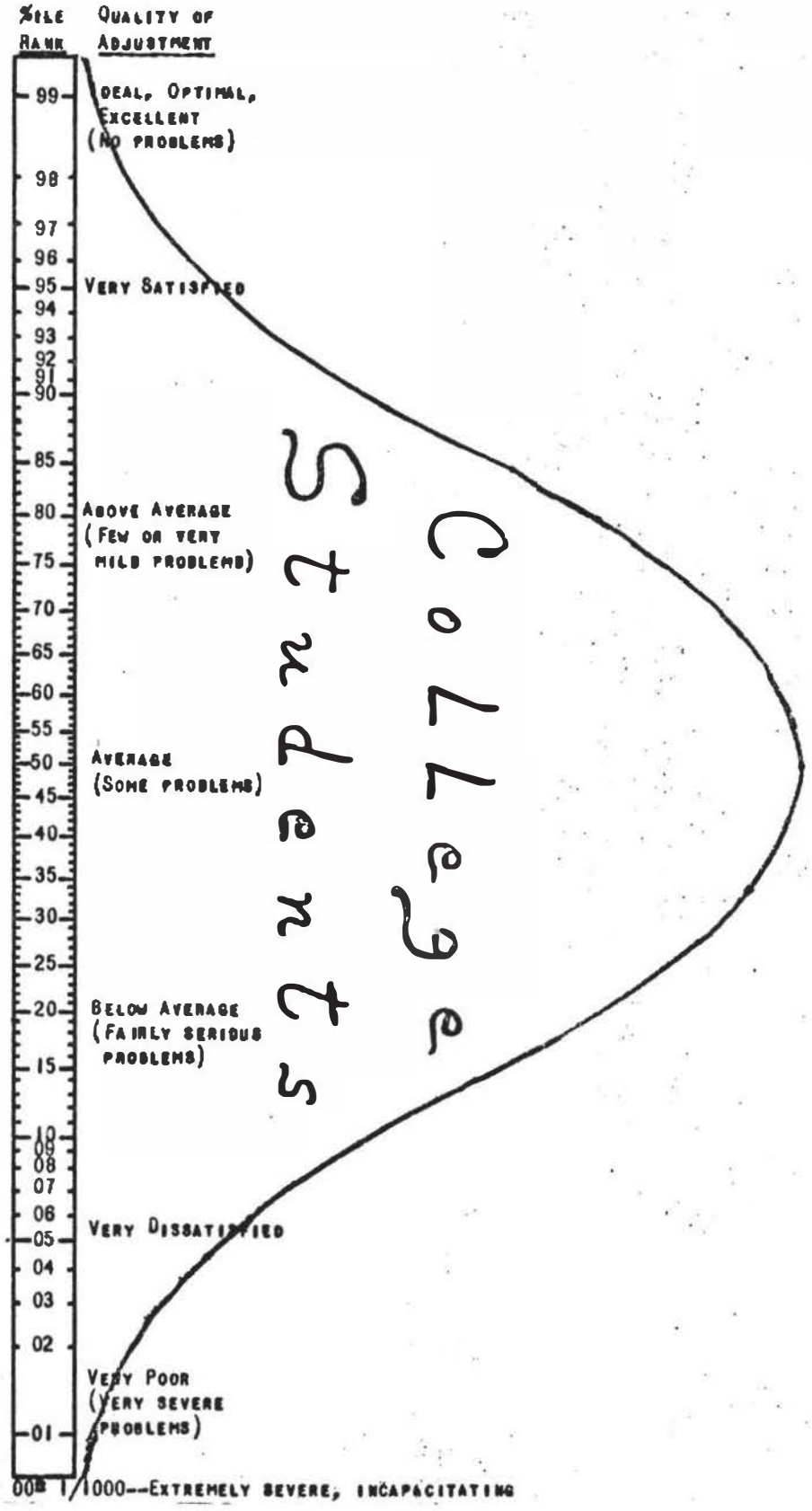
Scale 13: _____

Scale 14: _____

Scale 15: _____

Scale 16: _____

Scale 17: _____



RATINGS OF ADJUSTMENT -- INSTRUCTION SHEET

You will be given seventeen rating scales, which cover a wide variety of personal and interpersonal factors. The good or optimal end of each scale is described for you. Likewise, the poor or "problem" end of the scale is described. In the middle is the "typical" or average adjustment, i.e. some but not serious continuing problems.

The purpose of this instrument is to see how each person would rate him/herself in terms of his/her "level of adjustment" on a number of areas. There are no tricks or hidden meanings to any of these scales. We're only interested in what each person has to say about him/herself. Of course, the information will be kept confidential. Finally, we've noticed that people who answer the questions as honestly as possible find the ratings helpful in terms of giving them a chance to carefully consider just how they feel about themselves.

You should have the instruction sheet, a ratings booklet with descriptions, and an answer form with the rating scale to be used on it.

On the answer form please fill in the name, date, age, course, and sex data blanks. Then, in the numbered blank for each scale please put the number corresponding best with your present level of adjustment in that particular area. Use the number which best describes your level -- avoid using only the numbers printed on the scale -- but use any point from 0-100 which you feel best reflects you present level.

If you have any questions, please ask.

RATINGS OF ADJUSTMENT -- INSTRUCTION SHEET

You will be given seventeen rating scales, which cover a wide variety of personal and interpersonal factors. The good or optimal end of each scale is described for you. Likewise, the poor or "problem" end of the scale is described. In the middle is the "typical" or average adjustment, i.e. some but not serious continuing problems.

When you originally rated yourself in January it was in terms of your level of adjustment at that time. The purpose of this study is to determine how you have changed since the previous rating. Rather than our taking a simple difference score between the first ratings and the present ratings, we would like you to rate yourself this time in light of the amount of change you honestly think has occurred since January.

Therefore, we are showing you your original score sheet so that you can now rate yourself in terms of the change, if any, you have undergone since that original rating. For example, you may have scored yourself originally on Scale 2: Comfort at a level of 50. Now your feel that you have changed in the past few months so that you may have a 25% improvement. Indicate this on the present answer sheet by putting your level at 75%.

We also give you the option of changing your original score if you now feel that that score is wrong. For example, again on the Comfort scale you have changed immensely but you had originally scored yourself too high, say the 90% level, leaving little room to show a large change. To now show this change you may mark on the original scale (i.e. 2. ~~90~~ -- 60) by crossing out the original, putting a score you think would have been more accurate, and on the present answer sheet, write down you present level which indicates more accurately the amount of change (i.e. 2. 95). Thus, in this example you have shown a 35% change.

Keep in mind that the most important factor is the amount of change -- not the particular level. Try to show as accurately as possible how much you have changed, in either direction, since you were originally tested.

After you have finished all the ratings, if there were any items you actively worked on changing this semester, please indicate this by circling on the answer form the numbers of these items. Only 2 items need to be circled if there were many.

Again, we're only interested in what each person has to say about him/herself. Of course, all information will be kept strictly confidential.

If you have any questions or need assistance in the procedure, please ask.

Thank you.

RATINGS OF ADJUSTMENT -- INSTRUCTION SHEET

You will be given seventeen rating scales, which cover a wide variety of personal and interpersonal factors. The good or optimal end of each scale is described for you. Likewise, the poor or "problem" end of the scale is described. In the middle is the "typical" or average adjustment, i.e. some but not serious continuing problems.

In January, you rated your levels of adjustment at that time. The purpose of this study is to have you re-rate your level of adjustment for that time. In other words, you are now to rate yourself in terms of what level of adjustment you think you had back in January. Do not rate your present level, but instead rate yourself as to what level you had at the original testing.

Again, we're only interested in what each person has to say about him/herself. Of course, all information will be kept strictly confidential.

If you have any questions or need assistance in the procedure, please ask.

Thank you.

RATINGS OF ADJUSTMENT -- INSTRUCTION SHEET

You will be given seventeen rating scales, which cover a wide variety of personal and interpersonal factors. The good or optimal end of each scale is described for you. Likewise, the poor or "problem" end of the scale is described. In the middle is the "typical" or average adjustment, i.e. some but not serious continuing problems.

When you originally rated yourself in January, it was in terms of your level of adjustment at that time. The purpose of this study is to determine how you have changed since the previous rating. Rather than our taking a simple difference score between the first ratings and present ratings, we would like you to rate yourself this time in light of the amount of change you honestly think has occurred since January.

Therefore, we are showing you your original score sheet so that you can now rate yourself in terms of the changes, if any, you have undergone since that original rating. For example, you may have scored yourself originally on Scale 2: Comfort at a level of 50. Now you feel that you have changed in the past few months so that you may have a 25% improvement. Indicate this on the present answer sheet by putting your level at 75%.

Keep in mind that the most important factor is the <u>amount of change</u> -- not the particular level. Try to show as accurately as possible how much you have changed, in either direction, since you were originally tested.

After you have finished all the ratings, if there were any items you actively worked on changing this semester, please indicate this by circling on the answer form the numbers of these items. Only 2 items need to be circled if there were many.

Again, we're only interested in what each person has to say about him/herself. Of course, all information will be kept strictly confidential.

If you have any questions or need assistance in the procedure, please ask.

Thank you.

RATINGS OF ADJUSTMENT --
DESCRIPTION BOOKLET

Scale 1: Recognition of problems -- from aware to unaware.

-- Optimal adjustment -- Very aware of your problems and their causes; does not deny personal problems; not overly defensive with others.

-- Extremely severe -- Unaware of problems or their causes; denial of problems; overly defensive with others.

Scale 2: Comfort -- from relaxed to anxious.

-- Optimal adjustment -- Ideally calm; able to relax; easy going; content; confident.

-- Extremely severe -- Tense; uneasy; easily upset; overly concerned; worrying.

Scale 3: Happiness -- from very happy to depressed.

-- Optimal adjustment -- Enjoys life; cheerful; active; eager; optimistic; zestful living.

-- Extremely severe -- Depressed; sad; lonely; guilty; shameful; can't forgive self for some things in the past.

Scale 4: Productivity -- from high to low.

-- Optimal adjustment -- Fulfilling potentials; hard working; very productive.

-- Extremely severe -- Inactive; unproductive; not fulfilling potentials.

Scale 5: Emotionality -- from appropriate to inappropriate.

-- Optimal adjustment -- Not over or under-reacting; accepting of one's own feelings; free to experience wide range of emotions.

-- Extremely severe -- Over or under-reacting; inexplicable mood swings; repression of feelings.

Scale 6: Frustration tolerance -- from high to low.

-- Optimal adjustment -- Tolerates stress and anxiety well; "bears up" well, despite failures.

-- Extremely severe -- Falls apart under stress; gives up easily; inappropriate reactions to stress.

Scale 7: Self-understanding -- from very good to very poor.

-- Optimal adjustment -- Know strengths and weaknesses; aware of motives and feelings; aware of causes of behaviors.

-- Extremely severe -- Unaware of real feelings or their complexities; unrealistic about assets or weaknesses; uninsightful; unaware of the causes of behaviors.

Scale 8: Self-satisfaction -- from much to little.

-- Optimal adjustment -- Self-accepting; confident; good self concept.

-- Extremely severe -- Feels inferior or overly superior and arrogant; self-hatred or very critical of self; (Either extreme of superiority or inferiority are equally poor adjustments).

Scale 9: Physical appearance -- from completely satisfied to completely unsatisfied.

-- Optimal adjustment -- Satisfied with appearance; nothing you would change if you had a chance.

-- Extremely severe -- Totally displeased with appearance; feels ugly; feels deformed or awkward.

Scale 10: Sexual adjustment -- from satisfactory to unsatisfactory.

-- Optimal adjustment -- Satisfied with interests, attitudes, knowledge, and abilities.

-- Extremely severe -- Intense fears or inhibitions; disturbed by interests, impulses, or previous activities; lacking interest, knowledge, or ability.

Scale 11: Social skills -- from superb to poor.

-- Optimal adjustment -- Friendly; at ease; natural; comfortable; a good leader; gets along quite well with others.

-- Extremely severe -- Awkward; uneasy; introverted; withdrawn; can't get along with others.

Scale 12: Independence -- from optimal to low.

-- Optimal adjustment -- Self-reliant; self-directed; autonomous; can assume authority but can also be a good follower.

-- Extremely severe -- Very dependent; helpless; indecisive; overly submissive.

Scale 13: Love -- from satisfactory to lacking.

-- Optimal adjustment -- Able to love and be loved; feels appreciative and able to show it.

-- Extremely severe -- Feels rejected; insecure; jealous; untrusting; afraid of emotional involvement.

Scale 14: Acceptance of others -- from high to low.

-- Optimal adjustment -- Tolerant; empathetic; generally likes other people; respects others; accepts them as equals.

-- Extremely severe -- Hostile; sullen; intolerant; very critical; unsympathetic; envious.

Scale 15: Assertiveness -- from satisfactory to unsatisfactory.

-- Optimal adjustment -- Able to satisfy own needs without infringing on others; able to say "no" or make own desires known.

-- Extremely severe -- Overly passive; unable to express self appropriately; overly aggressive; "steps on" others.

Scale 16: Overall personal adjustment -- from excellent to poor.

Reflects adjustments to work, school, personal traits, etc. (Do not include interpersonal adjustment).

Scale 17: Overall interpersonal adjustment -- from excellent to poor.

Reflects how well you get along with others but not including your personal adjustment.

Appendix B

$$D = \sqrt{\Sigma(X-Y)^2}$$

Subject #1

Rating Items	Pre-test	Changed Original	Diff.	Post-test	Diff.	Diff.^{2}
1	80			95	-15	225
2	95			98	- 3	9
3	95			100	- 5	25
4	90			95	- 5	25
5	70			95	-25	625
6	97			98	- 1	1
7	95			95	0	0
8	90			99	- 9	81
9	90			90	0	0
10	90			99	- 9	81
11	99			99	0	0
12	90			99	- 9	81
13	99			99	0	0
14	95			97	- 2	4
15	80			99	-19	361
16	95			99	- 4	16
17	80			95	-15	225
						1759

$$D = 41.94$$

Appendix C

Subject	Original Rating	Memory Rating
1.	66	71
2.	75	82
3.	72	79
4.	77	88
5.	80	86
6.	83	90
7.	68	78
8.	87	90
9.	77	75
10.	51	55
11.	74	83
12.	83	89
13.	67	64
14.	69	79
15.	75	90
16.	70	78
17.	68	83
18.	89	91
19.	86	90
20.	84	85
21.	53	59
22.	70	71
23.	69	74
24.	82	94
25.	85	75
26.	85	80
27.	76	86
28.	69	88
29.	97	91
30.	93	93
31.	49	74

32.	71	83
33.	75	74
34.	74	78
35.	61	62
36.	86	86
37.	62	81
38.	65	73
39.	78	75
40.	80	82
41.	90	78

Appendix D

Descriptive Statistics

Subfile G1-1 N=29

Mean	61.483	Std. Err.	6.390	Std. Dev.	34.410
Var.	1184.044	Kurtosis	.549	Skewness	.910
Range	150.000	Minimum	10.000	Maximum	160.000

Subfile G1-2 N=23

Mean	66.348	Std. Err.	5.535	Std. Dev.	26.546
Var.	704.692	Kurtosis	.656	Skewness	.903
Range	109.000	Minimum	31.000	Maximum	140.000

Subfile G1-3 N=27

Mean	49.963	Std. Err.	4.679	Std. Dev.	24.311
Var.	591.037	Kurtosis	-.421	Skewness	.210
Range	95.000	Minimum	6.000	Maximum	101.000

Subfile G1-4 N=31

Mean	59.129	Std. Err.	6.149	Std. Dev.	34.237
Var.	1172.183	Kurtosis	-.267	Skewness	.758
Range	124.000	Minimum	16.000	Maximum	140.000

Subfile G2-1 N=20

Mean	48.950	Std. Err.	6.264	Std. Dev.	28.014
Var.	784.787	Kurtosis	.125	Skewness	1.012
Range	100.000	Minimum	18.000	Maximum	118.000

Subfile G2-2 N=15

Mean	65.867	Std. Err.	7.005	Std. Dev.	27.132
Var.	736.124	Kurtosis	.829	Skewness	1.129
Range	103.000	Minimum	31.000	Maximum	134.000

Subfile G2-3 N=14

Mean	67.214	Std. Err.	7.231	Std. Dev.	27.056
Var.	732.027	Kurtosis	-.605	Skewness	.476
Range	93.000	Minimum	29.000	Maximum	122.000

Subfile G2-4 N=16

Mean	53.625	Std. Err.	7.863	Std. Dev.	31.451
Var.	989.183	Kurtosis	-.562	Skewness	.632
Range	104.000	Minimum	15.000	Maximum	119.000

Statistics Using Dependent

D Scores from Procedure 2

Subfile G1-2 N=23

Mean	55.087	Std. Err.	6.263	Std. Dev.	31.035
Var.	902.083	Kurtosis	.963	Skewness	.845
Range	134.000	Minimum	7.000	Maximum	141.000

Subfile G2-2 N=15

Mean	53.400	Std. Err.	6.710	Std. Dev.	25.988
Var.	675.400	Kurtosis	-.870	Skewness	.600
Range	86.000	Minimum	18.000	Maximum	104.000