DOCUMENT RESUME

ED 256 981 CG 018 203

AUTHOR McNamara, Kathleen; Horan, John J.
TITLE Attending to Experimental Construct Validity in the Evaluation of Cognitive and Behavioral Treatments for Depression.
PUB DATE Apr 85
NOTE 34p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Attitude Change; *Behavior Modification; *Cognitive Restructuring; Comparative Analysis; *Counseling Effectiveness; Counseling Techniques; *Depression (Psychology); Reinforcement

ABSTRACT

Two current approaches to the conceptualization and treatment of depression have received considerable attention from the scientific community. The cognitive approach (Beck) posits that depression derives from negatively distorted beliefs that must be challenged in the context of cognitive therapy until they are replaced with positive and realistic thought patterns. The behavioral approach (Lewinsohn) views depression as a consequence of reinforcement deprivation, suggesting that treatment be directed toward increasing the frequency and variety of pleasure-producing activities. Clients (N=40) seeking service at a university counseling center were randomly assigned to one of four treatment conditions (cognitive, behavioral, combined, control). Pre- and post-test measures of depression included four cognitive measures, three behavioral scales, and two diagnostic inventories. Analysis of data revealed that the cognitive treatment factor produced a consistent and durable impact on devices reflecting cognitive manifestations of depression; some generalization to the behavioral domain occurred as well. The behavioral factors failed to produce improvement within the corresponding behavioral assessment battery or on any cognitive device. The obtained pattern of convergent and divergent outcomes indicated considerable construct-valid strength for cognitive therapy applied to a moderately depressed population. (Author/NRB)

Paper presented at the annual meeting of the

American Educational Research Association, Chicago, 1985

Attending to Experimental Construct Validity in the Evaluation of Cognitive and Behavioral Treatments for Depression

Kathleen McNamara

Colorado State University

and

John J. Horan

Division of Counseling and Educational Psychology

The Pennsylvania State University

RUNNING HEAD: Construct Validity in Depression Treatment

Requests for reprints should be sent to Kathleen McNamara, Department of Psychology, Colorado State University, Fort Collins, Colorado 80523. Copies of the treatment and/or assessment materials are also available for the costs of reproduction and postage; please specify the information requested.

Abstract

Distilled versions of cognitive (Beck) and behavioral (Lewinsohn) treatments for depression were crossed in a 2x2 design that included combined and high-demand control treatments as well. Multivariate and univariate analyses of pre-, mid-, post-, and follow-up data revealed that the cognitive treatment factor produced a consistent and durable impact on a battery of devices reflecting cognitive manifestations of depression; some generalization to the behavioral domain occurred as well. The behavioral factor failed to produce improvement within the corresponding behavioral assessment battery or on any cognitive device. Post-mortem analyses of a full syndrome measure suggested possible evidence favoring each factor. Both conditions generated equivalent demand characteristics and counselor ratings of client adherence to treatment. No interactions involving the treatments occurred. The obtained pattern of convergent and divergent outcomes indicates considerable construct-valid strength for cognitive therapy applied to a moderately depressed population. Possible reasons for behavior therapy's comparatively weak showing are discussed.

Attending to Experimental Construct Validity in the Evaluation of Cognitive and Behavioral Treatments for Depression

Depression was first described as a clinical syndrome by Hippocrates in the fourth century B.C. He and other ancients believed it to be caused by a superabundance of "black bile" in the brain. Other single-causal-agent hypotheses have been posited down through the ages including, for example, aggression turned inward (Fenichel, 1945) and deficiencies in endorphin production (c.f. Maier & Seligman, 1979; Romano & Turner, 1985).

Two current monolithic approaches to the conceptualization and treatment of depression have received considerable attention from the scientific community. The cognitive approach, most notably espoused by Beck (1967; 1974; Beck, Rush, Shaw, & Emery, 1978) posits that depression derives from negatively distorted beliefs that need to be subtly but persistently challenged in the context of cognitive therapy until they are replaced with positive and realistic thought patterns. The behavioral approach, on the other hand, views depression as a consequence of reinforcement deprivation. Lewinsohn (1974; Lewinsohn, Biglan, & Zeiss, 1976) thus suggests that treatment be directed toward increasing the frequency and variety of pleasure-producing activities. Social skills training is also recommended (so as to maximize the reinforcement obtained from others).

Seligman's (1975; Abramson, Seligman, & Teasdale, 1978) learned helplessness model and Rehm's (1977) self-control model also occupy

prominent places in contemporary literature. Both invoke cognitive and behavioral concepts to explain the etiology of depression. However, since treatment procedures derived from these two models are ultimately similar to those of Beck and Lewinsohn, we will not consider them further.

The phenomenon of competing schools--cognitive vs. behavioral--spawned three comparative studies (Taylor & Marshall, 1977; Shaw, 1977; and Hodgson, 1981) that as a group offered some evidence for a general treatment effect, but shed little light on the question of relative efficacy. In addition to their common methodological compromise of single-experimenter counselors, all but one dependent variable in this group of studies were global measures of depression which, in retrospect, cloud the issue of <u>how</u> the two treatments produced their equivalency (see also Hollon's (1981) discussion on "mechanisms of change" specified by a particular theory). This perplexing evaluation issue rarely receives attention outside the field of instructional psychology wherein Porter, Schmidt, Floden, and Freeman (1978), for example, pointed out the problem of using standardized achievement tests to assess the outcomes of competing arithmetic curricula. Total score comparisons ignore the fact that the individual items may differ in their relevance to the various interventions. Thus analyses of treatment-by-item interactions may be necessary to clarify the relationship between independent and dependent variables. It is interesting to note that in the Hodgson study the behavioral treatment did register an effect on the behavioral measure, but regrettably no specific cognitive measures were included that might have permitted a parallel finding. Moreover, Taylor and Marshall's reported superiority of the

combined treatment over the cognitive and behavioral interventions deployed alone raises the speculation that the greater efficacy derives from the possibility that each intervention impacted separate (but equivalent numbers) of treatment-relevant items on each global measure; combining interventions would thus produce the significantly higher total scores.

In addition to the discomforting ambiguities frequently produced by comparative studies employing global measures, our understanding of therapeutic outcomes is further muddled by the routine failure of our literature to address the issue of differential diagnosis (Horan, 1980). This problem is particularly acute in depression intervention in spite of the fact that many writers have assaulted the assumption that depression is homogeneous with regard to etiology and treatment responsiveness (e.g., Craighead, 1980; Hersen, 1981; Rush, 1982). Unfortunately, with the exception of bipolar depression (for which lithium is the treatment of choice), existing classification schemes (e.g., DSM III, RDC) have not proved useful in identifying those clients likely to benefit from a particular intervention mode (Rush, 1982). Although Craighead (1980) and Hersen (1981) have called for separating depressives on the basis of cognitive and behavioral skill deficits and then matching clients to treatment, we have found very few depressed clients whose deficits fall purely in one domain. Perhaps the issue of differential diagnosis could be more productively addressed by focusing instead on the specific and possibly unique effects of the various therapies.

Such attention would not only permit a more comprehensive cataloguing of a given treatment's effects, but the obtained outcome

pattern would also have strong implications for the construct validity of the experiment itself. Construct validity is usually associated with judgments on the value of various assessment devices. Campbell and Fiske's (1959) multitrait-multimethod matrix, for example, suggests that a measure is construct-valid to the extent that it: (a) correlates with different methods for measuring the same construct, and (b) fails to correlate with similar methods for measuring different constructs.

The construct validity of the counseling experiment likewise can be thought of in terms of convergent and discriminant relationships (see Cook & Campbell, 1979; Horan, 1984). In other words, do the manipulated variables produce theoretically consistent changes on measures which they are supposed to influence, and do they reliably fail to produce differences on theoretically unrelated variables? Whereas the Campbell and Fiske paradigm pays attention to the degree of correlation, our experimental simile focuses on the magnitude of the effect size.

Two experimental studies which bear directly on the question have failed to support the construct validity of depression treatment. Zeiss, Lewinsohn, and Munoz (1979) compared therapeutic regimens focusing on interpersonal skills, pleasant activities, or cognitions; and used outcome measures specifically keyed to those interventions. Although their study was methodologically commendable in many respects, low statistical power (6.8 Ss per comparison) may have been responsible for the differentially null effects. Relatedly, Wilson, Golden, and Charbonneau-Powis (1983) found durable wholesale gains produced by cognitive and behavior therapies vis-a-vis waiting-list controls on global measures of depression; however, no

differences appeared between the two active treatments on either the global measures or indeed on other measures specifically keyed to the cognitive and behavioral treatments. Again, low statistical power as well as differentially stringent comparisons (i.e., two active treatments set against each other rather than the no-treatment control) may have precluded the appearance of a construct-valid outcome pattern.

The findings of three other studies, however, appear quite promising. Rehm, Fuchs, Roth, Kornblith, and Romano (1979) reported that self-control therapy was more effective on global measures of depression and on a specific measure of self-control in comparison to assertion training, which in turn differentially impacted a measure of assertiveness. Given that self-control training is a broad cognitive-behavioral package and that assertion training addresses only a very small subset of the depression syndrome, the Rehm et al. study is perhaps best viewed as a comparison of a comprehensive program with an attention- or minimal-treatment control condition. Its construct validity implications nonetheless remain quite intriguing.

Similarly, Di Mascio, Weissman, Prusoff, Neu, Zwilling, and Klerman (1979) examined the separate and combined effects of pharmacotherapy and a form of psychotherapy that included both cognitive and behavioral strategies. The differential impact was most interesting. Pharmacotherapy's effects were initially on the vegetative symptoms such as sleep disturbance, somatic complaints, and loss of appetite. Psychotherapy, in contrast, registered early changes on mood, suicidal ideation, work, interests, and guilt.

Finally, McKnight, Nelson, Hayes, and Jarrett (1984) isolated nine subjects from a pool of 72 volunteers for depression treatment; three had deficits in social skills, three had irrational cognitions, and three had both kinds of dysfunction. A combination alternating-treatment-multiple-baseline design revealed that the effectiveness of cognitive and behavior therapy depended upon the relevance of either treatment to the assessed clinical problem, exactly what one would expect if the mechanisms of change postulated by cognitive and behavior therapies are valid. The small n and highly selective screening criteria, however, attenuate the generalization potential of these findings.

Our own study was designed with two purposes in mind. First, we were attempting to explore further the construct validity of depression treatment; in other words, do cognitive and behavior therapies actually impact their intended process and outcome measures while failing to produce, for example, differences in demand and degree of implementation? Second, we hoped to ascertain the relative power of each approach and the cost-benefit wisdom of combining them by noting if a given treatment produces positive changes on depression measures in addition to those of high theoretical relevance. (Such "crossovers" unaccompanied by internally consistent effects would render suspect the theoretical basis of the intervention.) Evidence for treatment efficacy coupled with a theoretically consistent outcome pattern (i.e., one with appropriate convergent and divergent effects) would have pronounced clinical implications as well: for example, treatments known to impact specific measures would be deemed

appropriate for clients presenting deficiencies on those particular components of an assessment battery.

Method

Subjects

Clients seeking services at the counseling center of a large southwestern university were accepted for this study if they (a) reported during an intake interview a depressive episode of at least two weeks duration, (b) produced a Beck Depression Inventory score $\geq$ 18 at intake and $\geq$ 16 at pretest, (c) obtained a combined score $\geq$ 20 on a modified version of the Hamilton Rating Scale for Depression applied to a videotaped pretreatment diagnostic interview, (d) presented no clinical evidence of suicidal behavior, psychosis, drug addiction, sociopathy, organicity, and/or major medical illness, and (e) gave informed consent to participate.

Subject recruitment continued over the course of a full academic year until 40 clients completed treatment; tested clients who did not meet the inclusion criteria (n = 13) or who declined to participate (n = 4) were referred to other counseling center services. Slight nondifferential attrition occurred; of 50 clients who initially qualified, 10 withdrew during the course of treatment (ns = 2 or 3 per cell). The final subject pool was largely female (73%), unmarried (85%), and young (Mean = 23, range 19-31).

Counselors

Seven doctoral interns in clinical and counseling psychology and one master's level social worker (4 M and 4 F) served as counselors. At the time of recruitment, seven counselors were self-described as "cognitive-

behavioral" in orientation; the eighth preferred the term "interpersonal" (cf. Strong, 1968). All counselors had expressed complete willingness to follow the exact procedures required by this study, despite any idiosyncratic preferences that might occur. Each counselor received approximately eight hours of didactic instruction, modeling of treatment procedures, roleplaying practice, and performance feedback prior to seeing any clients. They were also provided with treatment manuals and closely monitored (via audiotapes) for adherence to the appropriate intervention throughout the course of the study.

Assignment Procedures

Whenever four clients met the screening criteria, they were randomly assigned without exception to one of the four treatment conditions. Such "flights" of four were added to the subject pool over the course of two semesters and a summer session; a 4 x 3 chi square analysis indicated that treatment was adequately balanced over time ($\chi^2 = 4.07$ $p < .67$). Clients were also randomly assigned to counselors within the administratively imposed constraints that (a) total caseloads (ranging from 4 to 8) would reflect varying amounts of release time, and (b) gaps could not exist in a given counselor's schedule. Each counselor's caseload across the four treatments was perfectly balanced at pretest; a 4 x 8 chi square analysis conducted on posttest ns indicated that this equivalence endured throughout the study ($\chi^2 = 11.07$, $p < .96$). Numerous counselors, individually administered treatments, and small caseload-per-treatment ns were deliberately employed to preclude the possibility of counselor effects

interacting with treatment, to reduce mono-operation and mono-method biases (Cook & Campbell, 1979), and to enhance external validity.

Measures

The measures of depression used in this study fall into three categories, namely cognitive, behavioral, and diagnostic/generalization. The cognitive cluster was composed of the following:

1. The Automatic Thoughts Questionnaire (ATQ), is a 30-item self-report device, developed and validated by Hollon and Kendall (1980), which measures the frequency of negative thoughts associated with depression.

2. The Cognitive Scale (CS), is a 15-item Likert-type instrument designed specifically for this study to assess the extent to which clients learned and adopted the cognitive skills they were taught in treatment. Sample items are: Even when someone is unfriendly for no good reason, I automatically think it is my fault; I give myself pep talks when I feel discouraged or pessimistic. Since the CS was directly keyed to the treatment manual, it might be construed as an independent variable manipulation check as well as a measure of outcome gain; pretest internal consistency was found to be .68 using Cronbach's (1951) coefficient alpha.

3. The Recalled Cognitions (RC) exercise involved independent judges rating the quality of the client's thinking. Each client was asked to participate in a videotaped ten-minute getting-acquainted exercise with one of four research assistants who had been instructed to let the client initiate and maintain the conversation, but to respond in a friendly manner. The clients then watched their videotapes under instructions to "relive" the experience, that is, to recall in as much detail as possible all thoughts,

images, and feelings during the interaction. The clients were permitted to stop the videotape at any time in order to provide full descriptions. Audiotapes of these descriptions were independently judged on the number of negative, unpleasant, or disparaging statements made by the client. Pretest and posttest interrater reliabilities were .89 and .92 respectively.

4. The Self-Evaluated Social Skills (SESS) rating is composed of the same items as the Observer-Evaluated Social Skills (OESS) rating described below. Essentially, the clients evaluated their performance during the foregoing social interaction task; Lewinsohn et al. (1980) report internal consistencies of .89 and .91 for self-rated applications of this device.

The behavioral cluster was composed of the following:

1. The Pleasant Events Schedule (PES, MacPhillamy & Lewinsohn, 1971, later modified by Lewinsohn & Graf, 1973), is a 49-item self-report questionnaire which samples a broad range of potentially pleasurable activities. A crossproduct score reflecting the total amount of reinforcement obtained by the individual is derived by combining "frequency of occurrence" and "potential enjoyability" ratings for each item.

2. The Behavioral Scale (BS), is a 15-item Likert-type device (theoretically analogous to the CS described above). It was designed to assess the extent to which clients learned and adopted skills proffered by the behavioral treatment; pretest internal consistency was found to be .73. Sample items are: When I feel lonely or unhappy, I will call a friend and suggest an activity; I do not know what to say to people even though I want to talk to them.

3. The Observer-Evaluated Social Skills (OESS) rating was applied to the clients' performances in the social interaction task. As in Lewinsohn et al. (1980), videotapes of the clients were independently judged on 16 attributes of desirable social behavior. Inter-rater reliability coefficients were .92 at pretest and .98 at posttest. Lewinsohn et al. (1980) have reported internal consistency figures of .95 and .97.

Finally, the diagnostic/generalization category was composed of the following:

1. The Beck Depression Inventory (BDI), is a widely used 21-item self-report measure of overall depression level; its reliability and validity are well documented (see Beck & Beamesderfer, 1974). Cutoff scores of $\geq$ 10 and $\geq$ 16 respectively indicate mild and moderate depression; however, for establishing a stable diagnosis of depression in research work, clients need to obtain successive scores $\geq$ 16 on pretreatment testing occasions separated by at least two weeks (Hammen, 1980). Such was true in this study; moreover, to reduce the salience of a regression artifact, pretest scores were obtained from the second BDI administration.

2. The Hamilton Rating Scale for Depression (HRSD, Hamilton, 1960) lent strength to a pretreatment diagnosis of depression. Client responses in a diagnostic interview were videotaped and rated by two independent judges who obtained an interrater reliability coefficient of .88. As per scoring difficulties noted by Shaw (1977), three symptom categories were excluded from consideration by the judges (genital, hypochondriasis, and loss of insight). A combined cutoff score of $\geq$ 20 was required to

confirm the BDI diagnosis of at least moderate depression and to qualify for this study.

The measures pertinent to any given testing occasion were deployed in random order. All of the foregoing devices except the HRSD were administered both at pretest and again at posttest ten weeks later. A midpoint assessment occurred in the fourth or fifth week of treatment and a mailed follow-up evaluation took place two months after treatment ended. The midpoint and follow-up batteries included the ATQ and the CS from the cognitive category, the PES and the BS from the behavioral category, and the BDI.

In addition to the foregoing measures of depression, a battery of experimental-demand measures was employed:

1. Expectancy was assessed, as per recommendations by Borkovec and Nau (1972), at the end of the first session and again at midpoint by having the clients rate on 10-point scales five items pertaining to the logic of the treatment and their belief that the treatment would be successful.

2. Client Satisfaction with counselor and treatment was assessed posttreatment by readministering the expectancy questionnaire with modifications to verb tenses and certain adjectives.

3. Adherence was determined by the counselors who rated five posttreatment questionnaire items pertaining to the clients' completion of homework assignments, receptivity to suggestions, etc.

Treatment Procedures

Approximately ten days after their intake interviews at the counseling center and initial qualifying score on the BDI, all clients were given a one-

hour semi-structured diagnostic interview and asked to complete the pretest assessment battery. Those clients who continued to manifest depression on the BDI (later substantiated by the HRSD) and who expressed willingness to participate were then randomly assigned to one of three active treatments or to a high-demand control condition. All clients began treatment within ten days after the diagnostic interview (Mean = 6 days), and all treatments were individually administered in weekly audiotaped sessions of 50 minutes duration. The cognitive, behavioral, and control conditions required eight counseling sessions, the combined treatment 10. Perceived clinical necessity mandated an extra session or two for one client in each active treatment condition. Specific procedures were as follows:

1. The Cognitive Therapy (CT) condition was operationally keyed to the writings of Beck (1974, 1976; et al., 1979); however, in order to maximize procedural differences from behavior therapy, no attempts were made to modify the clients' behaviors or environments. Essentially, clients were told there is substantial research indicating that depression results primarily from the ways we evaluate our experiences rather than from unpleasant events per se. After being shown in great detail how feelings and behaviors are largely a function of thinking, clients were taught how to replace invalid assumptions and negative thoughts with constructive and realistic cognitions. Specific strategies included, for example, the recording of automatic thoughts and images, and identification of distortions, and various focused discussions about assumptions, beliefs, and attitudes relating to depression.

2. The Behavior Therapy (BT) condition was derived from the work of Lewinsohn and his associates (Lewinsohn, 1975; Lewinsohn & Graf, 1973; Lewinsohn & Grosscup, 1978; Lewinsohn, et al., 1976; Steinmetz, Antonuccio, Bond, McKay, Brown, & Lewinsohn, 1979); again, however, to avoid overlapping with cognitive therapy, no reference was made to cognitions as possible sources of depression. Essentially, clients were told there is substantiai research indicating that depression results from insufficient positive reinforcement. After being shown how improvements in mood state relate to increased participation in positively reinforcing (social and/or solitary) activities, clients self-monitored their activity levels as a precursor for later analyses and the learning of alternative behaviors incompatible with depression. Such behaviors included, for example, assertion, conversation, and social initiation skills fostered by modeling and roleplays. The clients were also cued and socially reinforced for increasing the frequency and variety of enjoyable activities in their daily lives.

3. The Combined condition included the rationales and strategies of both cognitive and behavior therapy. Essentially, clients were told that depression develops in two equally important ways, and consequently they were encouraged to modify both cognitive and overt behavior. Two extra sessions were found in pilot work to be sufficient for covering slightly abbreviated versions of the didactic material and homework assignments of the single treatment conditions. Possible rival hypotheses pertaining to differential contact time were a priori scheduled to be examined as in West, Horan, and Games (1984); however, the obtained outcome pattern described below obviated this need.

4. The High Demand control (C) condition provided the clients with a rationale and general goal statement (e.g., "to have a place to openly express feelings and have someone fully listen to you"). To achieve consistency, all counselors were instructed to adhere carefully to Chapter 2 of Rogers (1951).

Although we believe that elementary Rogerian relationship qualities (a) are beneficial for obtaining assessment data, (b) enhance the client's perception of the counselor as a powerful role model and source of secondary reinforcement, and (c) provide the clinical practice foundation from which cognitive and behavioral interventions are typically deployed (Horan, 1979), we are less optimistic about their exclusive theoretical relevance to depression treatment. Given recent ethical discussions about the use of placebo controls (e.g., Hodgson, 1981; O'Leary & Borkovec, 1978), our original design plan called for a "supportive" minimal-contact, informed, control condition, chained to eventual full treatment (or immediate intervention if clinically necessary). Since our hypotheses primarily concerned the differential impact of various treatments on specific measures, a high-demand, theoretically-irrelevant control condition, though desirable, was not methodologically necessary as long as the three active treatments generated equivalent demands. Oddly, however, counseling center policy mandated that we deploy Rogerian therapy in the control cell because (a) it was construed to be a strong and viable standard treatment, and (b) center clients were not permitted to receive any form of delayed or attenuated treatment. In effect, we found ourselves in the peculiar situation of being required to perform a more rigorous outcome evaluation

of cognitive and behavior therapy than we felt ethically comfortable in proposing, but which nonetheless had pronounced heuristic advantages.

Results

Preliminary Analyses

Pretreatment equivalence. One-factor ANOVAs conducted on pretest raw scores indicated that none of the four treatment conditions differed on any measure prior to treatment. Table 1 summarizes all treatment data on each testing occasion.

Insert Table 1 about here

Demand analyses. A one-factor ANOVA conducted on the expectancy measure administered after the first treatment session revealed no significant differences among the four treatments. A similar analysis of the midtreatment data yielded an overall effect [$F(3,28)=3.14$; $p < .04$], which when subjected to Scheffe post hoc comparisons showed higher expectations for success in the cognitive and behavior therapy conditions than in the combined and control conditions (i.e.,[(CT = BT) > (Combined = C)]. At posttest, however, no overall ANOVA effect appeared on the measure of client satisfaction. Placed in perspective, the foregoing pattern suggests that midway through treatment the CT and BT conditions were accompanied by raised expectations for improvement, but by the time of posttesting, the four conditions were again equivalent in perceived efficacy. Finally, a similar ANOVA conducted on the counselors' ratings of their clients'

adherence to treatment likewise indicated equivalence among the four conditions.

General Multivariate Outcome Analysis. A 2 x 2 (presence or absence of CT by presence or absence of BT) multivariate analysis of covariance (MANCOVA) was performed on the eight posttreatment outcome measures using pretreatment scores as covariates. (All slope assumptions were separately checked and met.) The Pillai-Bartlett Trace V criterion revealed a significant main effect in favor of cognitive therapy [$F(8,22) = 3.77$, $p < .006$]. No other multivariate main or interaction effects were found.

Specific Effects of Cognitive Therapy

Post hoc 2 x 2 univariate ANCOVAs on each posttest measure indicated that cognitive therapy produced significant or marginal main effects on the entire cognitive assessment battery: ATQ [$F(1,35) = 4.65$, $p < .04$]; CS [$F(1,35) = 3.71$, $p < .06$]; RC [$F(1,30) = 8.17$, $p < .007$]; and SESS [$F(1,30) = 3.99$, $p < .05$]. Cognitive therapy also produced a beneficial main effect on the OESS [$F(1,30) = 4.01$, $p < .05$], a rather stringent outcome criterion from the behavioral assessment battery. No interactions involving the cognitive and behavior therapy factors appeared on any measure.

Specific Effects of Behavior Therapy

The insignificant multivariate main effect for behavior therapy was, nevertheless, subjected to post hoc 2 x 2 univariate ANCOVA analyses in order to permit individual criterion comparisons with other published research. All for naught; the behavior therapy factor still failed to yield a significant main effect on any outcome device in either the behavioral or cognitive assessment batteries.

Post Mortem Analysis of the BDI

An initial post hoc 2 x 2 ANCOVA on the BDI generalization measure indicated no significant main effects or interactions involving the cognitive or behavior therapy factors. Visual inspection of the cell means, however, suggested the possibility of a floor effect and an exceedingly rigorous comparison condition. To be more specific, posttest means for the three active treatments clustered in the 4 to 6 range (on a scale of 1 to 63). The control condition, on the other hand, averaged very close to a cutoff score of 10, which by clinical and research convention, separates the "normal" from "mildly depressed" categories. Essentially then, at posttest the typical experimental client was "normal," and his/her control counterpart was very nearly so, having shown about two standard deviations of improvement.

The clients in each treatment and control condition were subsequently reclassified into two categories according to their posttest BDI scores. "Normals" scored 9 or below; all others, 10 or above. Fisher exact tests were then run on each of the three active treatment conditions in comparison to the control cell wherein four (out of nine) normals resided. The combined treatment cell, containing nine out of ten normals, produced a beneficial Fisher exact p of .049. Although the cognitive and behavior therapy conditions each had eight of ten normals, their respective contrasts with the control cell did not reach significance (ps = .129).

Midpoint and Two-Month Follow-Up Analyses

Parallel 2 x 2 ANCOVAs on the mid-point data using the pretests as covariates revealed no significant main effects or interactions on any measure. Apparently, the raised expectations produced by CT and BT on the

mid-point demand measure were not accompanied by actual therapeutic gain.

Similar 2 x 2 ANCOVAs on the follow-up data showed that despite the loss of power associated with further attrition, the posttest effects achieved by CT endured on the cognitive battery: ATQ [$F(1,21) = 4.23$, $p < .05$]; CS [$F(1,21) = 3.53$, $p < .08$]. A marginal effect favoring CT on the BDI was also now apparent [$F(1,20) = 3.21$, $p < .09$]. Again, however, none of the follow-up analyses supported the hypothesized efficacy of BT.

The phenomenon of continued attrition called for closer inspection (ns lost = 3, 4, 5, 6 for CT, BT, Combined, and C, respectively). We believe much of it can be attributed to the transient nature of student addresses, as six follow-up questionnaire packets were returned undelivered, and the addresses of five additional subjects were unknown. Nevertheless, two subsequent analyses ruled out the likelihood of attrition as an artifact. First, Fisher exact tests indicated equivalent follow-up dropout ns over all four treatment cells (CT x BT) and also between the two extremes (CT-C x Attrition-Retention). Second, a series of t tests comparing the posttest scores of cognitive therapy clients (i.e., CT and Combined) who completed the follow-up questionnaires with those who did not, revealed no differences on any measure.

Discussion

Our study attempted to evaluate the separate and combined effects of cognitive and behavioral therapies for depression in the context of experimental construct validity considerations. Cognitive therapy produced a theoretically consistent impact on an entire battery of devices reflecting

cognitive manifestations of depression. Its efficacy also generalized to an Observer Evaluated Social Skills (OESS) rating, a rather stringent criterion from the behavioral assessment battery. (If an effect on the OESS were to have appeared in the absence of any impact on the cognitive battery, the theoretical framework of cognitive therapy would have been open to question.)

A follow-up assessment battery, administered two months after treatment ended, indicated that the effects of the cognitive factor were durable. Moreover, since no beneficial changes were evident at a midtreatment assessment point, it would appear that the full cognitive program is a sine qua non for improvement.

All of these therapeutic gains occurred in the presence of equivalent experimental demands and adherence-to-treatment ratings. Thus, the obtained pattern of convergent and divergent outcomes indicates considerable construct valid strength—as well as treatment efficacy—for cognitive therapy applied to a moderately depressed population.

Behavior therapy's showing, on the other hand, was comparatively weak. It failed to produce any sign of improvement within a corresponding behavioral assessment battery; nor did it yield significant differences on any cognitive and/or follow-up measure.

Given the foregoing outcomes of the cognitive and behavioral factors, and the absence of statistical interactions between them on any dependent variable, cost-benefit considerations would seem to suggest deploying cognitive therapy alone regardless of the particular facets of an individual client's depression. After all, one might argue, cognitive therapy did all

that it was supposed to do—and more—while behavior therapy produced wholesale null effects and no incremental utility when combined with cognitive therapy. Such an indictment, however, may be premature, if not altogether unwarranted.

In the first place, the effects of cognitive therapy were, as expected, predominantly on the cognitive aspects of the depression syndrome. One cannot, having failed to impact the entire array of depressive behaviors, simply deny their clinical importance and redefine depression treatment as addressing mainly cognitive outcomes, at least not on the basis of the data at hand. Other factors may have been responsible for behavior therapy's comparatively impotent performance. For instance, we suspect that our behavioral assessment battery may be somewhat less sensitive than the cognitive battery in the detection of real changes in their respective psychological domains (or the outcomes themselves may be differentially malleable). The Pleasant Events Schedule, for example, is a frequent anchor in the behavioral assessment literature, yet to our knowledge it has never served to showcase differences between two competing treatments having equivalent demand characteristics. (Anecdotally, our counselors reported that their behavioral clients had indeed made noticeable progress in mastering behavioral skills, and they expressed surprise that such gains were not reflected in the data analysis.) Null results are as much a function of assessment adequacy as intervention efficacy.

Moreover, we must also be open to the possibility that despite our training efforts and a priori judgments of equivalent counselor competence in all treatment cells, in retrospect our counselors may have been

differentially more proficient in cognitive therapy. (Such superiority would not necessarily show up in the demand and adherence analyses.) We are especially vexed by the failure to register an effect on the Behavioral Scale. Recall that the items in this device were directly keyed to the behavioral treatment; thus lack of significant differences here could possibly be construed as a failure to manipulate the independent variable rather than an indicator of treatment ineffectiveness. Related to the issue of counselor proficiency is our additional post hoc observation that the focii of cognitive therapy appear more circumscribed—or at least easier to manage—than the diverse criteria for behavioral improvement. Simply put, cognitive therapy may be easier to do.

Finally, our post-mortem analysis of the Beck Depression Inventory (BDI) provides a slightly broader perspective from which to view the efficacy of behavior therapy. Given its prominence in the literature as a depression criterion, and its inclusion of items pertaining to somatic complaints (as well as cognitive and behavioral dysfunction), we classified the BDI in our diagnostic/generalization category. At posttest we noted that 8 of the 10 clients who received either cognitive therapy or behavior therapy alone, and 9 of the 10 clients who received both, displayed BDI scores in the normal range. Thus, according to the most widely employed criterion of clinical depression, behavior therapy would also have to be judged as very successful. We would not disagree. However, in the context of construct validity considerations, the theoretical mechanisms by which behavior therapy achieved its BDI efficacy have yet to be confirmed.

References

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. L. (1978). Learned helplessness in humans: Critique and reformulation. Journal of Abnormal Psychology, 87, 49-74.

Beck, A. T. (1967). Depression: Causes and treatment. Philadelphia: University of Pennsylvania Press.

Beck, A. T. (1974). The development of depression: A cognitive model. In R. J. Friedman & M. M. Katz (Eds.), The psychology of depression: Contemporary theory and research. New York: Winston-Riley.

Beck, A. T. (1976). Cognitive therapy and the emotional disorders. New York: International Universities Press.

Beck, A. T., & Beamesderfer, A. (1974). Assessment of depression, the Depression Inventory. In P. Pichot (Ed.), Psychological measurements in psychopharmacology and modern pharmacopsychiatry (Vol. 7). Basal, Switzerland: Kerger.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). Cognitive therapy of depression. New York: The Guilford Press.

Borkovec, T. D., & Nau, S. D. (1972). Credibility of analogue therapy rationales. Journal of Behavior Therapy and Experimental Psychiatry, 3, 257-260.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Craighead, W. E. (1980). Away from a unitary model of depression. Behavior Therapy, 11, 122-128.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally College Publishing.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

DiMascio, A., Weissman, M., Prusoff, B. A., Neu, C., Zwilling, M., & Klerman, G. L. (1979). Differential symptom reduction by drugs and psychotherapy in acute depression. Archives of General Psychiatry, 36, 1450-1456.

Fenichel, O. (1945). The psychoanalytic theory of neurosis. New York: Norton.

Hamilton, M. (1960). A rating scale for depression. Journal of Neurology, Neurosurgery, and Psychiatry, 23, 56-62.

Hammen, C. L. (1980). Depression in college students: Beyond the Beck Depression Inventory. Journal of Counseling Psychology, 48, 126-128.

Hersen, M. (1981). Complex problems require complex solutions. Behavior Therapy, 12, 15-29.

Hodgson, J. W. (1981). Cognitive versus behavioral-interpersonal approaches to the group treatment of depressed college students. Journal of Counseling Psychology, 28, 243-249.

Hollon, S. D. (1981). Comparisons and combinations with alternative approaches. In L. P. Rehm (Ed.), Behavior therapy for depression. New York: Academic Press.

Hollon, S. D., & Kendall, P. C. (1980). Cognitive self-statements in depression: Development of an Automatic Thoughts Questionnaire. Cognitive Therapy and Research, 4, 383-395.

Horan, J. J. (1979). Counseling for effective decision-making: A cognitive-behavioral perspective. Monterey, CA: Brooks/Cole.

Horan, J. J. (1980, December). Experimentation in counseling and psychotherapy. Part I: New myths about old realities. Educational Researcher, 5-10.

Horan, J. J. (1984). On the construct validity of experiments in counseling and psychotherapy. New Fellow Address at the Annual Meeting of the American Psychological Association, Toronto, Canada.

Lewinsohn, P. M. (1974). A behavioral approach to depression. In R. J. Friedman & M. M. Katz (Eds.), The psychology of depression: Contemporary theory and research. New York: Winston-Wiley.

Lewinsohn, P. M. (1975). Engagement in pleasant activities and depression. Journal of Abnormal Psychology, 84, 729-731.

Lewinsohn, P. M., Biglan, A., & Zeiss, A. M. (1976). Behavioral treatment of depression. In P. O. Davidson (Ed.), The behavioral management of anxiety, depression and pain. New York: Brunner/Mazel.

Lewinsohn, P. M., & Graf, M. (1973). Pleasant activities and depression. Journal of Consulting and Clinical Psychology, 41, 261-268.

Lewinsohn, P. M., Mischel, W., Chaplin, W., & Barton, R. (1980). Social competence and depression: The role of illusory self-perceptions? Journal of Abnormal Psychology, 89, 203-212.

Lewinsohn, P. M., & Grosscup, S. J. (1978). Decreasing unpleasant events and increasing pleasant events: A treatment manual for depression. Unpublished mimeograph, University of Oregon.

MacPhillamy, D. J., & Lewinsohn, P. M. (1971). A scale for the measurement of positive reinforcement. Unpublished mimeograph, University of Oregon.

Maier, S. F., & Seligman, M. E. P. (1979). Learned helplessness: All of us were right (and wrong): Inescapable shock has multiple effects. In G. H. Bower (Ed.), Psychology of learning and motivation. New York: Academic Press.

McKnight, D. L., Nelson, R. D., Hayes, S. C., & Jarrett, R. B. (1984). Importance of treating individually assessed response classes in the amelioration of depression. Behavior Therapy, 15, 315-335.

O'Leary, K. D., & Borkovec, T. D. (1978). Conceptual, methodological, and ethical problems of placebo groups in psychotherapy research. American Psychologist, 33, 821-830.

Porter, A. C., Schmidt, W. H., Floden, F. E., & Freeman, D. J. (1978). Practical significance in program evaluation. American Educational Research Journal, 15, 529-539.

Rehm, L. P. (1977). A self-control model of depression. Behavior Therapy, 8, 787-804.

Rehm, L. P., Fuchs, C. Z., Roth, D. M., Kornblith, S. J., & Romano, J. M. (1979). A comparison of self-control and assertion skills treatments of depression. Behavior Therapy, 10, 429-442.

Rogers, C. R. (1951). Client-centered therapy. Boston: Houghton-Mifflin.

Romano, J. M., & Turner, J. A. (1985). Chronic pain and depression: Does the evidence support a relationship? Psychological Bulletin, 97, 18-34.

Rush, A. J. (Ed.). (1982). Short-term psychotherapies for depression. New York: The Guilford Press.

Seligman, M. E. P. (1975). Helplessness: On depression, development, and death. San Francisco: Freeman.

Shaw, B. F. (1977). Comparison of cognitive therapy and behaviaor therapy in the treatment of depression. Journal of Consulting and Clinical Psychology, 4, 543-551.

Steinmetz, J., Antonuccio, D., Bond, M., McKay, G., Brown, R., & Lewinsohn, P. M. (1979). Instructor's manual: Coping with depression. Unpublished mimeograph, University of Oregon.

Strong, S. R. (1968). Counseling: An interpersonal influence process. Journal of Counseling Psychology, 15, 215-224.

Taylor, F. G., & Marshall, W. L. (1975). Experimental analysis of a cognitive-behavioral therapy for depression. Cognitive Therapy and Research, 1, 59-72.

West, D. J., Horan, J. J., & Games, P. A. (1984). Component analysis of occupational stress inoculation applied to registered nurses in an acute care hospital setting. Journal of Counseling Psychology, 31, 209-218.

Wilson, P. H., Goldin, J. C., & Charbonneau-Powis, M. (1983). Comparative efficacy of behavioral and cognitive treatments of depression. Cognitive Therapy and Research, 7, 111-124.

Zeiss, A. M., Lewinsohn, P. M., & Munoz, R. F. (1979). Nonspecific improvement effects in depression using interpersonal skills training, pleasant activities or cognitive training. Journal of Consulting and Clinical Psychology, 47, 427-439.

Table 1

Means and Standard Deviations Produced by Each Treatment on All Dependent Measures at Each Testing Occasion

		Treatment Condition			
		Cognitive Therapy	Behavior Therapy	Combined Therapy	High Demand Control
		Automatic Thoughts Questionnaire			
Pretest	M	99.30	97.30	93.90	102.80
	SD	22.06	9.81	17.19	21.42
Midtest	M	77.50	75.55	82.00	71.57
	SD	17.24	16.87	8.33	34.21
Posttest	M	52.90	62.90	48.40	56.90
	SD	5.35	4.80	4.08	13.43
Followup	M	49.58	63.33	51.00	60.80
	SD	12.15	17.73	10.23	17.48
		Cognitive Scale			
Pretest	M	49.40	48.70	49.20	52.40
	SD	4.98	6.77	3.79	5.40
Midtest	M	44.60	46.80	44.00	47.28
	SD	4.90	6.56	4.20	8.63
Posttest	M	39.80	42.50	38.80	44.90
	SD	4.42	8.95	7.21	5.90
Followup	M	38.71	39.83	37.00	43.80
	SD	3.73	5.78	6.38	2.17
		Recalled Cognitions			
Pretest	M	7.70	8.80	8.05	8.35
	SD	1.66	3.22	2.77	2.10
Posttest	M	5.20	6.50	5.25	7.50
	SD	1.34	2.79	1.91	1.92

Table 1
(continued)

		Treatment Condition			
		Cognitive Therapy	Behavior Therapy	Combined Therapy	High Demand Control
		Self-Evaluated Social Skills			
Pretest	M	63.60	64.30	54.30	60.20
	SD	13.44	8.84	8.68	11.01
Posttest	M	69.90	67.87	62.50	63.00
	SD	15.87	4.61	9.84	12.67
		Pleasant Events Schedule			
Pretest	M	1.66	1.25	1.52	1.52
	SD	.53	.35	.58	.69
Midtest	M	1.99	1.48	1.82	1.61
	SD	.25	.43	.31	.24
Posttest	M	2.07	1.76	2.12	2.04
	SD	.16	.55	.45	.48
Followup	M	2.08	1.76	1.92	1.89
	SD	.17	.63	.49	.63
		Behavioral Scale			
Pretest	M	46.70	51.70	50.11	47.00
	SD	8.00	5.60	6.68	5.66
Midtest	M	44.89	45.78	46.16	46.00
	SD	6.53	6.28	2.32	9.76
Posttest	M	40.60	43.10	42.90	44.50
	SD	5.52	5.44	5.97	5.25
Followup	M	40.71	40.83	42.00	44.00
	SD	2.50	3.97	7.79	4.18

Table 1
(continued)

Observer Evaluated Social Skills					
Pretest	M	68.25	67.00	61.65	66.50
	SD	11.08	7.70	9.13	9.72
Posttest	M	71.65	69.12	64.50	66.77
	SD	9.25	4.73	8.08	10.02
Beck Depression Inventory					
Screening	M	26.67	30.22	23.50	27.20
	SD	4.18	5.99	5.42	6.94
Pretest	M	24.80	25.90	22.11	25.55
	SD	5.29	4.04	4.28	8.35
Midtest	M	16.20	15.44	14.57	16.86
	SD	5.35	4.80	4.08	13.43
Posttest	M	6.50	5.50	4.80	9.67
	SD	4.17	3.56	3.55	5.75
Followup	M	4.71	6.17	4.75	8.60
	SD	1.70	2.23	1.89	3.21