DIFFERENTIAL EFFECTS OF RELAXATION TECHNIQUES ON TRAIT ANXIETY: A META-ANALYSIS

KENNETH R. EPPLEY

Stanford University

ALLAN I. ABRAMS

Foster City, California

JONATHAN SHEAR

Virginia Commonwealth University

Hand and computer searches located studies on the effects of relaxation techniques on trait anxiety. Effect sizes for the different treatments (e.g., Progressive Relaxation, EMG Biofeedback, various forms of meditation, etc.) were calculated. Most of the treatments produced similar effect sizes except that Transcendental Meditation had significantly larger effect size ($p < .005$), and meditation that involved concentration had significantly smaller effect. Correlations with effect size were calculated for many variables, e.g., population, age, sex, experimental design, duration and hours of treatment, pretest anxiety, demand characteristics, experimenter attitude, type of publication, attrition, etc. Only a few variables (mainly population, duration, hours, and attrition) significantly influenced effect size. Controlling for possible confounding variables did not alter the overall conclusions. The difference in effect size between treatments was maintained both when only published studies were included and when only the studies with the strongest design were included. Possible explanations for the findings are examined.

Hundreds of studies in the literature have examined the effects of a variety of relaxation techniques in reducing the effects of stress and improving physical and mental health. While positive results often have been reported, it has not been established conclusively whether any of these procedures are more effective than others or whether,

---

Portions of this paper were presented by the third author at the August 1984 convention of the American Psychological Association in Toronto, Canada. The authors wish to thank Arthur Aron for many useful suggestions. We thank Michael Dillbeck for assistance with the literature search.

Correspondence that concerns this article should be addressed to the first author at Bin 26, SLAC, PO Box 4349, Stanford University, Stanford, CA 94309.

indeed, any of them produce larger effects than placebos. There has been little systematic effort to integrate the literature in such a way that objective conclusions could be drawn. Narrative (i.e., non-statistical) reviews of the literature often arrive at quite different conclusions based on much the same data. A quantitative meta-analysis, however, has the advantage of being able to integrate the data from a larger number of studies. It has the potential to display significant quantitative relationships that narrative reviews can overlook. It also has the potential to resolve apparent disagreements among studies when these disagreements are due to differing statistical power, different experimental conditions, etc.

The present study examines the effects of various relaxation techniques on a single outcome measure, trait anxiety. A large segment of society suffers from high trait anxiety, which is associated with significant health and behavioral problems (Sarason & Spielberger, 1975).

The authors were particularly interested in comparing the effectiveness of various somatic techniques, such as muscular relaxation, and various cognitive techniques, such as meditation. (There are, of course, procedures that combine both.) Therefore, the study was designed to provide a clear comparison between these modalities.

Method

Literature Search

Dependent variable. The topic most frequently examined in studies on relaxation was trait anxiety, and this became the dependent variable of the present study. "Trait" anxiety refers to the general tendency to be anxious as distinguished from "state" anxiety, the degree of anxiety at a particular moment. Trait measures should be more reliable in the sense that they are more stable over time. Although they are self-report, these instruments, particularly the Spielberger State-Trait Anxiety Inventory (STAI) which was used most often, are well validated in hundreds of studies (Spielberger, Gorsuch, & Lushene, 1970). That is, large correlations exist between anxiety as measured by these instruments and as measured by clinical and behavioral evaluations. The test-retest and between-instrument correlations for the trait anxiety instruments are among the highest found for psychological measures. There are also experimental construct validations in which manipulations that would be expected to alter anxiety do in fact show the expected results on state measures.

Previous meta-analyses such as those of Smith and Glass (1977) commonly have examined a number of measures rather than a single one, as in the present study. Often, differences between measures have been considerably larger than differences between treatments. The present authors were interested primarily in possible treatment differences and felt that the use of a single outcome measure would greatly reduce possible confounding effects. As we noted, trait anxiety instruments are among the most reliable of the self-report measures. While it also would have been interesting to examine behavioral measures, such data were extremely sparse for relaxation treatments. Data on physiological measures were more extensive; however, these were generally quantities (e.g., skin resistance, metabolic rate, etc.) that display very large within-subject variation. That is, they are basically state rather than trait measures. Thus, their interpretation would be more ambiguous and more subject to possible confounding effects.

Search implementation. The database for this study was obtained through hand and computer searchers of the literature.[1] All entries in *Psychological Abstracts, Science Citation Index, Social Sciences Citation Index, Comprehensive Dissertation Index, Medline, Excerpta Medica, and Sociological Abstracts,* from the earliest entries through

[1]A complete list of studies used in the meta-analysis is available from the first author on request.

1982, were examined by computer search, using the keywords "relaxation" and "meditation." Bibliographies of papers were searched by hand.

The number of studies found for Progressive Relaxation (PR) was far larger than for the other treatment categories, and it was possible to obtain adequate statistical power using only studies published in journals. To increase the statistical power of the other categories, their database was extended to include Ph.D. dissertations, M.A. theses, and research anthologies. It was determined that there was no statistically significant difference between the results obtained using journal studies alone and those obtained using this expanded database. (See below.)

To keep the search and analysis manageable, only studies that involved the two most common treatments, PR and Electromyographic (EMG) Biofeedback, were included in the relaxation category. Thus, a number of treatments that had only a few studies each were not included unless they occurred as alternate treatments in the other studies used.

There were many relaxation studies that involved psychiatric patients, but such populations were not found commonly in the meditation database. A preliminary analysis indicated that patients tended to show larger effects than other subjects. Both to reduce the number of studies to a more manageable size and to prevent a possible systematic bias, we did not include relaxation studies that used psychiatric patients unless these studies already were included from the meditation category.[2]

The studies included in the meta-analysis satisfied the following conditions: (a) The treatments were identifiable as forms of relaxation or meditation; (b) the instruments were identifiable as trait anxiety measures; (c) sufficient data were reported to calculate effect sizes; (d) the study employed either pre-post design or post-test with random assignment; and (e) the subjects were not experienced in relaxation or meditation techniques at pretest.

The search located 110 studies published in journals that involved relaxation and anxiety. Of these, 13 could not be obtained, and 21 did not use trait anxiety measures. Of those that remained, 40 involved PR. Of these PR studies, 8 did not adequately report data, 8 were excluded because they used psychiatric patients, 2 were excluded for design reasons (i.e., they did not satisfy criteria (d) or (e) above), and the remaining 22 PR studies were included. Of 113 studies that involved mediation and anxiety, 3 could not be obtained, 22 did not use trait anxiety measures, 12 failed to report data adequately, 6 were excluded for design reasons, and the remaining 70 were included. Seventeen studies on EMG Biofeedback were selected in a similar manner.

Coding validation. All coding was done by the first author. An independent coding validation was performed by a colleague (a Ph.D. in psychology) on 10 of the studies chosen randomly. Agreement between coders was 86%.

Treatment categories. Treatments were divided into four major categories, each of which contained a similar number of outcomes: (a) Progressive Muscular Relaxation (PR). The majority of research on relaxation treatments has involved PR. Therefore, it was put into a separate category. (b) Other forms of relaxation (OR). All other relaxation techniques were placed in this category. (c) Transcendental Meditation (TM). This is a highly standardized mental procedure introduced by Maharishi Mahesh Yogi, and it has been used in the majority of meditation studies. Therefore, it was put into a separate category. (d) Other forms of meditation (OM). All other techniques either described as meditation or similar to well-known forms of mediation were included in this category.

---

[2]These studies with patient populations were used only in the initial overall comparison, but were excluded for all groups in the final unconfounded comparisons. Thus, this selection should have produced no bias toward any treatment. Because relaxation often was used as an alternate treatment, the number of outcomes with patient populations was similar for the different treatments even in the overall comparison.

Those subcategories of OR and OM with at least a moderate number of effect sizes were examined separately. These were EMG Biofeedback, Benson's "Relaxation Response," concentration forms of meditation, meditation that used Sanskrit mantras and a permissive mental attitude (these techniques model the general features of TM), and placebos. This last category included any treatment described as a placebo provided that it involved some actual mental or physical "relaxation" procedure.[3]

Subject variables. The subject populations in the studies were entered under the categories of college, high school, adult, psychiatric or drug abuse patients, children, adult prisoners, juvenile offenders, or elderly.[4] Screening of subjects (i.e., including only subjects with a minimum trait anxiety score) also was coded as categorical (nominal) data. Age, fraction of males, and pretest anxiety level were entered as continuous (metric) variables.

Subsidiary treatment variables. Also coded were the instruction mode (written, taped, or live, and individual or group); the duration of the study; the hours of instruction and/or meetings during the first week; hours of follow-up meetings during subsequent weeks; and assigned frequency of treatment sessions. Demand characteristics included rationale and degree of suggestion, payment of fees, whether teaching was done in a traditional context and/or with spiritual overtones, requirements to abstain from drugs, and whether the teaching was done by an organization that might have facilitated social contacts.

Therapist variables. Experimenter allegiance, therapist experience, and whether testing was administered by the therapist were coded as categorical data.

Outcome measures. The particular trait anxiety instruments used (e.g., STAI, TMAS) were entered as categorical data.

Design quality and internal validity. The classifications of design were: subjects used as own controls, non-random assignment to groups, random assignment, and random assignment with low attrition (less than or equal to 15%). Also coded were the number of subjects on pretest and posttest and the subject attrition.

Sources. Studies were classified by whether they were obtained from scientific journals, Ph.D. dissertations, or other sources (i.e., masters' theses, research anthologies, etc.).

Calculation of effect sizes. Following Glass (1977), we defined effect size as

$$ES = (X_{C2} - X_{E2})/SD \qquad (1)$$

X_{C2} and X_{E2} are the posttest scores for the control and experimental groups.

ES is an ideal effect size if there are no differences between groups on the pretest. This would be true for studies with random assignment to groups, a large sample, and small attrition. However, in our database there were a substantial number of studies with non-random assignment; these could produce a systematic bias on ES. Even in studies with random assignment and low attrition the sample sizes tended to be small, allowing in some cases substantial differences between groups on the pretest. In more than half of these studies, this difference was more than .25 standard deviations. While this should not produce a systematic bias, it could result in a significantly larger variance. Therefore, to adjust for differences on the pretest, we defined an alternate effect size:

$$ES_{adj} = (X_{C2} - X_{E2})/SD - (X_{C1} - X_{E1})/SD \qquad (2)$$

[3]We note that placebos were not included in the major categories or in any of the analyses except in comparisons to other treatments in Figure 1 and Table 5.

[4]Categories for which there was no clear criterion for ordering the entries, e.g., population, test instrument, etc., were coded as categorical (nominal) data.

If the experimental and control groups are equal on the pretest ES_{adj} reduces identically to ES. As predicted, the variance of ES was significantly larger than that of ES_{adj}.[5]

To determine whether ES or ES_{adj} was a more reliable measure, we examined those studies in which the experimental and control groups were similar on the pretest. For these studies ES and ES_{adj} agreed closely, as would be expected.[6] The mean value of ES_{adj} calculated for all studies was very similar to the mean effect size for those studies with groups similar on the pretest. ES did not agree as closely, particularly for studies that did not employ random assignment. Therefore, for brevity we report here only the results from ES_{adj}.

The above definition of effect size requires a no treatment group. Some studies used only alternate treatment groups as controls. To be able to include the outcomes from such studies we defined a third effect size, ES_{pp}:

$$ES_{pp} = (X_{E1} - X_{E2})/SD \tag{3}$$

The main utility in calculating ES_{pp} is for comparisons with a smaller number of effect sizes, in which the power might otherwise be inadequate. Because the results from ES_{pp} generally followed closely those from ES_{adj}, we will not report them here except in a few cases in which comparisons with ES_{pp} having a larger number of effect sizes confirmed a trend found in ES_{adj}.

Analysis

In the initial overall comparison, effect sizes for PR, OR, TM, and OM were compared with all outcomes and with all independent outcomes. Comparisons also were made between the two standardized treatments, PR and TM, and subcategories of OR and OM. Because the inclusion of weak data frequently is criticized in meta-analysis, the overall comparisons were repeated, first, including only the studies of highest validity, and, second, including only studies published in journals. A comparison was also made of the internal validity of studies drawn from the various sources.

Univariate analysis was performed on the continuous (metric) and categorical (nominal) variables to determine which were correlated significantly with effect size.[7]

---

[5]The standard deviations used were the control groups' posttest scores if reported. (A detailed description of how standard deviations were computed when data were not reported is available from the author.) In our definition of ES_{adj}, we use this single standard deviation in dividing both pre- and posttest scores. One might define an alternate adjusted effect size as:

$$ES'_{adj} = (X_{C2} - X_{E2})/SD_2 - (X_{C1} - X_{E1})/SD_1 \tag{2}$$

i.e., using the standard deviations from the pre- and posttest with the corresponding differences rather than a single standard deviation. However, in practice, there was little difference between the values of ES_{adj} and ES'_{adj}, and it would not have altered the conclusions to use ES'_{adj}. Therefore, we do not report results for ES'_{adj}.

Our definition of adjusted effect size is the simplest one that has these features: (a) it reduces identically to the standard ES when the groups are the same on the pretest, and (b) it is zero if the changes in the control groups and the treatment groups are the same.

From a formal point of view, it may appear reasonable that when one takes the difference between pre- and posttest values, one then should divide by the standard deviation of the difference scores. However, there are several practical reasons for preferring the definition that we chose. First, the SD of the differences is not commonly reported. Second, an effect size using this SD would not satisfy either (a) or (b) above and would not be directly comparable to the standard ES, even when groups were similar on the pretest. The disagreement is especially pronounced when there is a high correlation between pretest and posttest scores. Then, the standard deviation of the differences approaches zero, and using it would produce a large effect size, even when the standard ES is almost zero. Therefore, the definition we have used is more conservative.

[6]With this restriction the comparisons between treatments yielded the same results with ES as those reported with ES_{adj}. Results for ES and ES_{pp} are available on request.

[7]Categorical variables were analyzed by disaggregating the effect sizes into the different categories for that variable and examining the resulting groups for significant differences.

Possible confounds were identified as those variables that had significant correlations with effect size and differed between the treatment groups. A non-confounded comparison was made by matching on categorical variables and performing a multiple linear regression analysis on the continuous variables. (Another option would have been to use a general linear model ANOVA that included both continuous and categorical data. Because we had adequate statistical power, we opted to use the conceptually simpler matching method.) Because curvilinear effects conceivably might produce spurious results in the multiple linear regression, a comparison also was made by matching potential confounding variables. Finally, an analysis was made of controlled comparisons between different treatments in the same study.

RESULTS

Overall Effect Size

The overall effect sizes for the different treatment groups are given in Table 1.[8] The first comparisons was between PR, OR, TM, and OM.

Table 1
Effect Sizes: All Independent Outcomes

Treatment	Effect size	*SD*	*n*
Progressive relaxation (PR)	.38	.40	30
Other relaxation (OR)	.40	.35	37
All relaxation (AR)	.39	.37	67
Transcendental meditation (TM)	.70	.40	35
Other meditation (OM)	.28	.34	44

TM had significantly larger effect size than OM, PR, and OR (TM vs. OM, $t[77] = 4.95, p < .001$; TM vs. PR, $t[63] = 3.22, p < .005$; TM vs. OR, $t[70] = 3.38, p < .005$). All confidence levels are based on two-tailed tests.) The differences between the other groups were not significant.[9]

Each of the subcategories of OR and OM was compared to the two standard treatments, PR and TM (Figure 1). TM was found to have significantly larger effect size than all other groups. The comparisons between PR and other treatments were not significant except that PR was significantly larger than concentration meditation and borderline larger compared to Sanskrit mantras with permissive attitude.

Correlates of Effect Size

Tables 2 and 3 show the mean values and correlations with effect size for the continuous background variables for the different treatment groups.[10]

[8]The comparisons that used all outcomes (including repeated measures) gave results very similar to those with all independent outcomes. (In the case of repeated measures, independent outcomes were defined as the outcome with the longest duration or the most commonly used instrument.) Therefore, we report only results for independent outcomes.

[9]Because there was negligible difference between PR and OR, we will not report further results from OR. However, we have included results from All Relaxation (AR) because it represented a larger sample of studies.

[10]Correlations from all studies were generally similar to those from the studies matched on populations. Because the latter are less confounded and were used in the final analysis, these values are reported here.

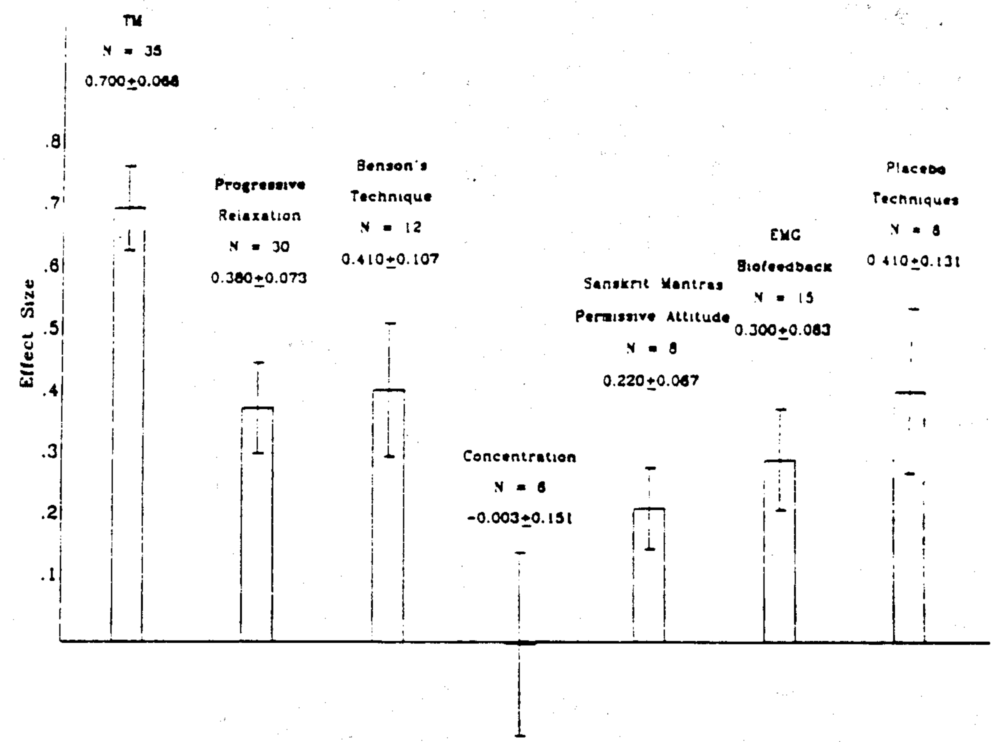

Fig. 1. Effect sizes for subcategories of meditation and relaxation, mean ± standard error.

Table 2

Background Variables for the Different Treatment Groups Studies with Matched Populations[a] Means (SD,n)

Variable	PR	AR	TM	OM
Duration (months)	1.6(1.1,28)	1.4(.96,61)	2.5(1.3,31)	1.8(1.2,39)
Age	21.(4.5,26)	23.(6.6,53)	25.(7.1,31)	24.(6.7,35)
Sex (males/total)	.42(.16,21)	.46(.22,48)	.51(.18,24)	.48(.19,32)
Hours-first week	1.3(1.0,24)	1.5(1.2,56)	5.5(0.0,24)	1.5(.92,34)
Follow-up hours	3.3(2.7,25)	3.7(2.6,57)	1.6(.54,24)	3.3(4.0,35)
Frequency of sessions (per week)	12.(3.8,18)	8.8(4.7,44)	14.(0.,31)	11.(3.9,36)
Number (pretest)	21.(11.,27)	18.(9.3,59)	28.(40.,26)	24.(13.,37)
Number (posttest)	16.(8.6,25)	16.(8.3,49)	22.(30.,31)	16.(8.8,33)
Attrition	.20(.19,24)	.19(.20,47)	.16(.18,26)	.32(.27,31)
Pretest score (STAI only)	43.(5.6,19)	41.(8.0,43)	41.(4.3,17)	39.(8.5,29)

[a]High school, college, adult.

Table 3
Correlations with Effect Sizes: r(n)
Studies with Matched Populations

Variable	PR	AR	TM	OM
Duration (months)	−.03(24)	.17(49)	.36(26)*	.29(37)*
Age	−.25(20)	.06(43)	−.36(23)*	.35(28)*
Sex (males/total)	−.46(15)*	−.19(32)	−.02(17)	−.22(24)
Hours-first week	−.24(19)	−.20(41)		−.23(26)
Follow-up hours	.47(20)**	.49(42)†	.52(16)**	.38(27)**
Frequency of sessions (per week)	.08(15)	−.07(31)		.06(28)
Number (pretest)	−.22(22)	−.17(44)	.26(19)	.07(30)
Number (posttest)	−.54(19)**	−.33(38)**	.07(23)	.03(26)
Attrition	.33(19)	.31(37)*	.46(19)**	.20(25)
Pretest score (STAI only)	.14(25)	.24(33)	.47(15)*	.45(27)**

*$p < .1$. **$p < .05$. ***$p < .01$. †$p < .001$.

Subject variables. For TM, the studies with adult populations had significantly smaller effect size than those with high-school/college or prison, which did not differ significantly. For OM, PR, and All Relaxation (AR), adult was not significantly different than high school/college. For all non-TM treatments, psychiatric patients showed significantly larger ES_{pp} when compared to all non-patient groups.

Populations that were screened to have high anxiety showed significantly larger effect sizes than unscreened populations. There was also a significant correlation between pretest anxiety level and effect size. There was a borderline negative correlation between age and effect size for TM, but a borderline positive correlation for OM. There was a significant negative correlation between fraction of males and effect size.

Subsidiary treatment variables. Effect size was correlated significantly with follow-up hours and duration of treatment. (The correlation with duration was borderline on ES_{adj} but significant on ES_{pp}.) Instruction mode, hours during the first week, and assigned frequency of treatment sessions were not correlated significantly with effect size. No demand characteristics (e.g., rationale, payment of fees, etc.) were correlated significantly with effect size except that the non-TM treatments taught by an organization had smaller effect size than treatments not so taught.[11]

We also looked for possible non-linear interactions between instruction modes, i.e., combining live plus individual instruction, or live, individual instruction plus a high degree of suggestion. However, there was no indication of a larger effect due to either of these combinations.

Therapist variables. Therapist variables were not correlated significantly to effect size.

Outcome measures. The effect sizes for different test instruments did not differ significantly.

Design quality and internal validity. The differences between different experimental designs were not significant except that studies with random assignment had borderline

[11]Because there was an indication that college populations showed larger effects than adult, conceivably there could have been an expectation effect in the college subjects that was masked in the overall comparison. However, when college populations were examined as a block, there were no significant differences for any demand characteristics.

larger effect size than non-random assignment for OM. As a test, the overall comparisons between treatments were repeated including only the studies of highest internal validity, with random assignment to groups and low attrition. The TM studies still had significantly larger effect sizes than the other treatments.

There was a significant correlation between attrition and effect size. There was a significant negative correlation between number on the posttest and effect size.

Sources. There were no significant differences for any of the groups between effect sizes from journals, dissertations, or other sources. To confirm this, we repeated the overall comparisons between groups including only studies published in journals. The difference between TM and the other treatments was still significant.

It is sometimes suggested that dissertations are inherently less reliable than journal articles. There is no a priori reason for this to be so because dissertations must satisfy the requirements of a reviewing committee, which conceivably could be more stringent than those of a referee. Dissertations should meet certain minimum standards of quality because the experimental design typically must be approved by the committee. Smith and Glass (1977) found no significant differences on internal validity between studies from journals or dissertations.

It can be argued that including the dissertation outcomes was not only valid, but necessary. A number of authors (e.g., Hedges, 1984; Lane & Dunlap, 1978) have pointed out that effect sizes found in journals may not be an unbiased sample. That is, researchers are probably more likely to submit, and reviewers more likely to accept, studies that found significant effects as opposed to those that did not. Also, due to space limitations in journals, authors are less likely to report means for nonsignificant findings. In dissertations, the results are generally presented regardless of whether or not they were significant. Also, the greater space allows considerably more detail to be reported. One way to see whether this is an important effect is to compare the results from journals to data found in dissertations and presumably less subject to such influences. This strategy was followed by Smith and Glass (1977).[12]

We also examined the internal validity characteristics for the different sources. We tabulated the number of studies without a separate control group, with non-randomly assigned controls, and with random assignment. We compared the number with and without alternate treatment groups and also the sample size and the attrition rate. Significantly more of the studies obtained from books used subjects as their own controls. There were no significant differences between journal and dissertation studies on method of assignment. The dissertation studies were more likely to use alternate treatments and had larger sample size, but both differences were of borderline significance. The only factor in which the dissertation studies had lower validity than the journal studies was in having a significantly higher attrition rate. However, when we adjusted for the difference in attrition, there was still no significant difference between the effect sizes. The larger attrition for the non-TM dissertations would not confound the comparison between the other treatments and TM because it would only tend to make the non-TM effect sizes larger (since there was a positive correlation beween effect size and attrition). The TM dissertations did not have a larger attrition than the TM journal studies.

While the weaker design of the studies obtained from books was not associated with any differences in effect size, to be conservative we repeated all the analyses both including and excluding these studies. The results remained the same.

---

[12]Hedges (1984) suggests that one may adjust the effect sizes based on the sample size to control for this effect. However, because the TM studies had a slightly larger sample size, this would have only tended to increase the differences found.

In the various comparisons above, the fact that there may have been no main effect between conditions (e.g., between studies from dissertations vs. journals) does not absolutely prove that there were no differences, even when the power was adequate. Conceivably, differences between populations or other parameters could have acted as moderator variables and, thereby, have masked out possible differences between conditions. To look for such effects, we reexamined those comparisons that involved potential confounding variables between TM and the other treatments and selected studies with college populations as a block (that being the largest group). Similarly, we examined those parameters that were correlated significantly with effect size to see whether they could have acted as confounding moderating variables. (To give a concrete example, we looked for any significant differences between duration, attrition, hours of treatment, etc., between dissertation and journal studies.) In only a few cases were there significant differences between parameters with the right sign to potentially confound the results, and in none of these cases did adjusting for this change the conclusions.

Identification of Possible Confounding Variables

The possible confounding categorical variables were population and screening. TM had more studies with prison populations, and AR had more studies with patients and studies that screened subjects for high anxiety.

For the continuous variables, the TM studies had significantly greater duration than OM or AR. TM had significantly more hours during the first week, but significantly fewer hours of follow-up. OM had significantly greater attrition than either TM or AR, which were similar. TM also had borderline larger number on the pretest. There were no other significant differences between the groups on potential confounding variables.

The only continuous variables for which there was a significant difference between the treatments and also a significant correlation with effect size were duration, attrition, and follow-up hours.

Unconfounded Comparisons

Multiple regression analysis. The groups were matched on population by selecting only studies with high school, college, and adult populations, unscreened for high anxiety. A multiple linear regression was performed, adjusting for duration, attrition, and follow-up hours (Figure 2).[13] The differences between TM and the other treatments were all significant and accounted for between 29% and 41% of the total variance. (TM vs. OM, $t[52] = 5.93, p < .001$; TM vs. PR, $t[43] = 4.74, p < .001$; TM vs. AR, $t[66] = 5.25, p < .001$.) Differences between the other groups were not significant.

Matched comparison. Because multiple linear regression sometimes can produce spurious results (for example, when there are important curvilinear effects), to be conservative we also performed a comparison by matching on possible confounding variables. From the studies matched on population, a further selection was made from the OM, PR, and AR groups of studies with duration at least 6 weeks (Table 4). For these studies, all groups were similar on duration. Correcting for attrition and follow-up hours would only have tended to increase the difference between TM and the other groups. For this sample, there was, thus, no significant difference on any variable that should have produced a larger TM effect.

For the comparison between treatments with the studies matched for population and duration, TM had significantly larger effect size. The other groups did not differ significantly. The results were still significant if one included only studies from journals or dissertations.

---

[13]Ideally, one would like to do a multiple regression that included all factors that might be correlated with effect size. However, for N independent variables, this involves solving for N(N-1)/2 unknown partial correlations. Thus, the total number of effect sizes available was too small to include all of the possible correlates.

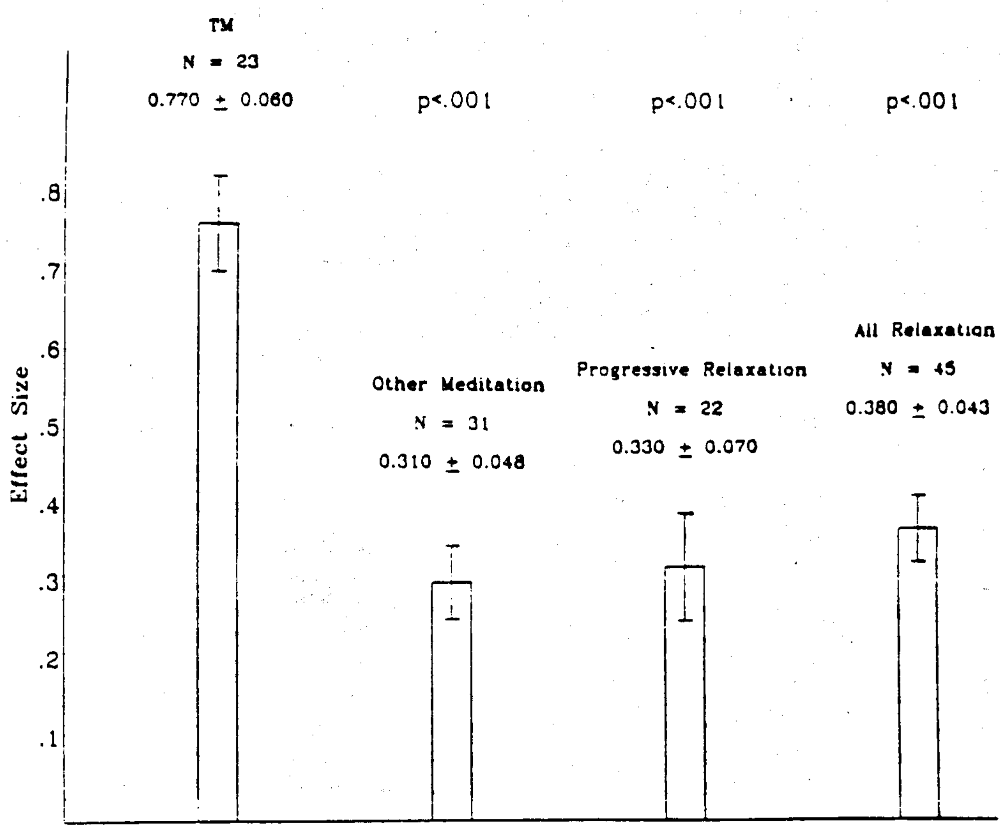

FIG. 2. Effect sizes, studies matched for population, adjusted for duration, attrition, and follow-up hours. The p values are for paired comparisons between TM and the other treatments.

Table 4
Studies Matched for Population and Duration[a]
Effect Size (SD,n)

Database	PR	AR	TM	OM
All matched studies	.35(.38,11)	.40(.33,16)	.73(.33,23)	.36(.30,17)
Journals/dissertations	.38(.39,10)	.43(.33,15)	.72(.33,17)	.36(.30,17)
Journals/dissertations	.20(.37, 5)	.24(.31, 7)	.77(.17, 6)	.22(.30, 4)
Random, low attrition				
Journals/dissertations	.36(.41, 9)	.41(.37,12)	.89(.37, 4)	.37(.32,14)
Random assignment, alternate treatment[b]				
Authors neutral or negative toward TM				

[a]Duration > 6 weeks for OM, PR, AR.
[b]Effect size was defined using the no treatment controls.

In the regression analysis, adjusting for the difference in hours of treatment between TM and the other groups only tended to increase the differences in effect size. To confirm this, a comparison was made between TM and the other treatments, matched for population, duration, and total hours of treatment. TM still had significantly larger effect size.

While we found no differences between live vs. taped instruction, there is some evidence (Paul & Trimble, 1970) that live instruction in PR produces a larger effect on physiological measures. However, all but two outcomes in the matched sample for PR involved live instruction, and excluding these two did not alter the conclusion.

The comparison was still significant if one restricted the studies, matched for population and duration, to those of highest validity, i.e., with random assignment and low attrition. We also reexamined the various correlations using these highest validity studies. The only difference with what had been found before was that the borderline positive correlation between duration and effect size became nonsignificantly negative. Because the difference between TM and the other groups was significant whether or not we controlled for duration, this would not affect the conclusion.

We also attempted to form a comparison between the "most reliable" outcomes. These were confined as studies matched for population and duration, found in journals or dissertations, with random assignment to alternate treatment groups, not done by researchers with a pro-TM allegiance. The differences between treatments were still significant.[14]

Controlled comparisons between treatments. We tabulated the differences in effect sizes between treatments directly compared within the same study (Table 5), for all comparisons of TM vs. OM, PR, or OR, PR vs. OM or OR, and placebos vs. OM, PR, or OR.[15] As one would expect, for these studies there were no significant differences on any population or pretest variables. The only differences between groups were that TM had borderline smaller attrition, which would not confound the comparison. Therefore, no further adjustment was performed. The differences between TM and the other groups were comparable to those found above and were statistically significant. The differences between PR and other treatments and placebos and other treatments were quite small and were not significant.

Table 5
Controlled Comparisons Between Treatments In the Same Study

Comparison	Difference in effect size	*SD*	*n*
TM vs. OM, PR, or OR	.38	.39	18
TM vs. OM, PR, or OR	.43	.40	15
Journal or dissertation			
PR vs. OM or OR	.03	.43	24
Placebo techniques vs. OM, PR, or OR	.03	.24	14

[14]Here the effect size was still defined using the no-treatment groups as controls. While the number of effect sizes was too small to include only those with low attrition, the comparison was still significant when we adjusted for attrition.

[15]There was only one study that compared TM to placebos, so we did not tabulate this category. The difference scores correspond to ES_{adj} with the substitution of the scores for the alternate treatment group in place of the control group.

Discussion

The analysis found that TM produced a larger effect size than OM, PR, or OR. This does not, however, prove that differences between the treatments were responsible for the result. For example, researchers such as Smith (1976) have described TM instruction as involving a high degree of suggestion and might argue that the larger effect size was mainly due to greater expectations on the part of the subjects. We wish to examine a number of such factors that might provide possible alternative explanations for the results.

Demand effects. There was some evidence that the amount of attention from the trainers did have an effect. There was a significant correlation between hours of follow-up treatment and effect size. However, the TM studies had fewer hours of follow-up than the other treatments. (The correlations were calculated controlling for treatment.) Conceivably, attention from the TM teachers could have been more effective due to their possibly greater experience. However, for other treatments, teachers with extensive experience did not produce larger effects than other teachers.

In TM training, almost all contact with the trainers is during the first 2 weeks, while contact typically was spread out over a longer period for the other treatments. If the difference between TM and other treatments were predominantly due to expectation and/or teachers' attention, this difference should have become smaller over time, or at least should not have increased. Yet, the difference in effect size between TM and all other treatments increased from .23 in the first month to .35 in the second month to .38 in subsequent months.

The meta-analysis also found no evidence that higher expectations resulted in larger effect size. There were no significant differences between treatments with minimal rationale and those with a high degree of suggestion. As a further test, we tabulated the findings of all those studies in our database that measured correlations between pretest expectations and subsequent measured change in trait anxiety. These correlations on average were quite small and only accounted for a few percent of the variance, not enough to explain the difference between TM and other treatments in the meta-analysis.

A number of studies (Abrams, 1977; Riddle, 1979; Smith, 1976; Throll, 1981; Zuroff & Schwarz, 1978) have attempted to equalize demand conditions between TM and alternate meditation and/or relaxation treatments. In these studies TM produced a significantly larger reduction in trait anxiety in 8 out of 10 such comparisons and was similar in effect in the other two. However, the interpretations drawn within the studies themselves by their authors were more mixed. We will discuss the two studies with the strongest design and the most careful matching of demand conditions, Zuroff and Schwarz (1978) and Smith (1976).

Zuroff and Schwarz (1978) attempted to duplicate all the demand characteristics of TM, using PR as the alternate treatment. Although they taught PR with a complex rationale that, according to measurement, produced pretreatment expectations comparable to TM, they found that TM produced a signficantly larger reduction in anxiety on one self-report instrument (the S-R Inventory). For the TM group, anxiety continued to diminish with time even though the expectations of the subjects decreased. However, a second self-report measure (the ACL) produced a treatment effect of only borderline significance. Results on a behavioral measure were not significant. The behavioral measure used was not correlated significantly with trait anxiety on the pretest and was not tested in any other study. Therefore, its validity and reliability are unknown. Thus, the overall results of this study cannot be regarded as clear-cut.

Smith (1976) compared TM to "Periodic Somatic Inactivity" (PSI, which he regarded as basically a placebo), to a mantra technique modelled on TM, and to a procedure of thinking as many positive thoughts as possible. He found that TM and PSI produced a larger reduction in anxiety than the other two treatments. He concluded that the difference was probably due to expectations because TM and PSI had more eleborate rationales. While this study was well designed, its conclusions cannot be interpreted entirely unambiguously. For example, Smith's study suffered from a high attrition rate in all groups. Further, it may not be valid to regard PSI as simply an intrinsically ineffective placebo, for the technique involved letting thoughts flow effortlessly coupled with the vague intention to sit quietly, and there are traditional schools of Zen and Yoga whose meditation instructions are quite simliar to PSI. Another difficulty in interpreting Smith's results arises from the fact that the TM group practiced more regularly than the PSI group, and Smith adjusted the scores to account for this. While this at first appears reasonable, it is possible that the TM subjects' tendency to practice more frequently was an effect of the treatment as well as a cause of it. That is, an ineffective treatment might well be expected to have more drop-outs than a highly effective one. Thus, Smith's results also must be regarded as inconclusive.

The most unambiguous test of the effect of expectations is to manipulate them experimentally while keeping other elements equal. None of the studies in our database that attempted this for relaxation and/or meditation techniques (e.g., Bassett, Blanchard, & Estes, 1977; Bradley & McCanne, 1981; Riddle, 1979) found significant differences on state and/or trait anxiety due to such manipulation, even when the manipulation was verified to be highly successful (Bassett et al., 1977).

The above observations appear to conflict with the common belief that expectation can produce large effects. While there have been experiments in which expectations produced unequivocal results, such as in physiological reactions to drugs (e.g., Luparello, Leist, Lourie, & Sweet, 1970), in these studies the measurements were performed immediately after the suggestions were made. In the studies on trait anxiety, the measurements were made weeks or months after the initial suggestion. It is generally held that effects of suggestion are not cumulative or long-lasting, and it has been observed that expectations with regard to meditation tend to decrease with time (Zuroff & Schwarz, 1978). Thus, in the trait anxiety studies any initial expectation effect may have declined to the point at which it was negligible compared to other factors.

To summarize, the meta-analysis found demand effects to be uncorrelated with effect size. Further, the difference between TM and the other groups increased with time, which should not occur as a result of expectations. Studies that have deliberately generated high subject expectations have not found this manipulation to produce a larger reduction in anxiety. Studies that compared TM to treatments with similar demand characteristics found a larger TM effect for the majority of such comparisons, although ambiguities remain in the interpretation of individual experiments. Thus, the weight of the evidence from our database does not appear to support the hypothesis that the larger TM effect was the result of suggestion or attention. Nevertheless, it would be highly desirable to have more well-designed experiments along the lines of Zuroff and Schwarz (1978), which might settle the matter conclusively.

Volunteer effect. There was no significant difference between the effect sizes found for studies in which subjects were assigned randomly to either TM or alternate treatments vs. those in which the subjects chose to learn TM. There were also no significant differences between studies with subjects who paid to learn TM and studies in which they learned without charge.

Regularity of practice. Greater regularity of practice by TM subjects may contribute to some degree to the larger effect size, but, if this were not due to initially stronger motivating factors, it would not discount the results. In general, the TM subjects do appear to have practiced more regularly (e.g., Smith, 1976; Zuroff & Schwarz, 1978). However, the median correlation coefficient between regularity and anxiety reduction in our database was only .17. This would account for about 3% of the total variance, which is too small to explain the difference between treatments.[16]

Unincluded variables. It is conceivable that important variables may have been overlooked in the coding or never reported in the primary studies. If such variables played a significant role in the overall comparisons between studies, they could be expected to play a lesser role in the comparisons within a single study. Yet the outcomes for these two groups were comparable.

Experimenter bias. Conceivably, the findings might be biased because some of the TM studies were done by experimenters with pro-TM allegiance. However, the effect sizes in studies done by persons with negative or neutral attitudes toward TM were slightly larger than those with pro-TM allegiance, and excluding the latter studies did not alter the conclusions.

Sampling bias. There were no significant differences between effect sizes from journals, dissertations, or books, and the TM studies maintained a significantly larger effect size if only journal articles were examined. Had there been some systematic bias toward suppressing TM studies with poor results, the TM effect sizes would not have been normally distributed, but should have been significantly skewed. However, this was not found.

Bias from inclusion of weak studies. There was no significant correlation between effect size and experimental design. TM maintained a significantly larger effect when only studies with the highest validity were included.

Differences between TM instruction and other treatments. The weight of the evidence from our database appears to suggest that the TM technique and/or the way in which it is taught (which are not in practice separable) produces larger effects than other meditation and relaxation procedures. The differences between TM and the other treatments found in the meta-analysis were robust and seem unlikely to be a statistical artifact.

However, it is less clear how to account for these differences. Those techniques that modelled the general features of TM failed to show a larger effect than other treatments. While TM instruction might have produced greater expectations than some of the other treatments, demand characteristics did not appear to account for the difference. TM training involved more hours during the first week than most of the other groups, but this variable was uncorrelated with effect size for the non-TM treatments.

The meta-analysis does, however, provide one possible clue to the source of the difference. The techniques that involved concentration produced a significantly smaller effect; in fact, the difference between relaxation and concentration was the only difference between treatments comparable to the difference between TM and relaxation. Thus, one factor might possibly be less effort involved in the practice of the TM technique. There is some experimental evidence that TM is more successful in bringing about a virtually effortless practice than other forms of meditation with generally similar features (Morse, Martin, Furst, & Dubin, 1977). TM instruction involves a great deal of time spent in ensuring that subjects meditate in a completely "effortless, spontaneous" manner. This involves, first, much emphasis on the point that any effort or concentration is detrimental to the practice of meditation, and, second, specific procedures ("checking") designed to lead the subject through the meditation process and to verify that he or she both understands and is gaining the experience of thinking a particular thought in a spontaneous manner. None of the other treatments in the meta-analysis duplicated either of these features.

In general, however, emphasis on the desirability of effortless relaxation does not appear to be sufficient in itself to bring about the state desired. It is clear that the intellectual knowledge that one ought to relax can fail to produce significant physiological

[16]We used assigned frequency of treatment/practice sessions rather than actual frequency of practice as a variable because only a handful of studies reported the actual frequency.

and psychological relaxation, as problems with hypertension show. Similarly, active attempts to relax and fall asleep are often counterproductive. Indeed, it might appear that the very attempt to produce deliberately a state of effortless relaxation generally will fail. Deliberate generation of such a state appears to require some process of thinking that is itself both intentional and at the same time spontaneous. Such a mode of thinking is, however, not impossible. There are situations in which the mind can become directed in a completely effortless manner, for example, when one attends to a piece of music or work of art that one particularly enjoys. This suggests that intentional mental activity, such as the mental repetition of a particular thought or sound, is likely to be most effortless if the activity itself possesses some intrinsically attractive quality. It is, therefore, not implausible that both the "mental object" and the mechanics used in a meditation technique should be relevant in this regard. TM teachers accordingly suggest that both the particular sound or thought (the "mantra") and the procedures used make an important difference to the process of meditation and to its results.

It is difficult to draw any conclusions from the meta-analysis concerning the matter of the "object" of meditation. Meditation techniques using Sanskrit mantras did not show a larger effect than other mantra techniques. However, the particular mantras used in these studies were chosen arbitrarily from a very large number of such sounds. Thus, this result is probably not a valid test of the hypothesis that there exist effective criteria for selecting appropriate sounds for different individuals. There is also some experimental evidence (Moltz, 1977; Morse, Martin, Furst, & Dubin, 1979) that meditation on different sounds does produce clearly different subjective experiences that vary from individual to individual. More research in this area could be useful.

There has been skepticism with regard to the claim by TM teachers that they can choose the optimal sounds for a particular individual, probably because no theory has been put forth to explain the connection between particular sounds and particular results. However, it is not entirely implausible that even rough, rule-of-thumb experimentation over centuries might produce more consistent empirical results than inventions by researchers based only on a few years of investigation.

Conclusions and Recommendations for Future Research

TM was found to produce a significantly larger effect than other forms of meditation and relaxation in the reduction of trait anxiety. This difference appeared too large to be accounted for by expectations or placebo effects. A comparison of the content of different treatments and their differential effects suggests that the difference was due, at least in part, to the greater degree of effortlessness of TM. However, the difference is of practical importance, regardless of the explanation. Even if factors such as greater experience by the teachers, or the weight of tradition, should turn out to play a larger role than was suggested by the data, it would not alter the "bottom line" finding that TM produced larger and more consistent effects.

An interesting question is whether the larger TM effect on trait anxiety also is found on other measures. Ferguson's (1981) study suggested that TM also may produce larger effects than other meditation treatments on other self-report instruments. A meta-analysis by Dillbeck and Orme-Johnson (1987) found a larger TM effect on a number of physiological measures, but only examined comparisons to unstructured relaxation. A tally of the results of all studies found in which TM was compared to other treatments gives some preliminary evidence of a larger TM effect on behavioral/performance measures and possibly on physiological measures.[17] More research in these areas is needed.

---

[17]Details available on request.

Obviously, the present study can draw conclusions only on techniques for which sufficient data existed in the literature. It is possible that there are procedures not examined in this meta-analysis that are equal to or superior to TM in reducing trait anxiety.

We recommend that future studies on relaxation and meditation ensure an adequate sample size, allowing for attrition. It would be valuable to attempt to measure the subjects' expectations and correlate them with outcome. It would be useful to report the subjects' regularity of practice, preferably in some standard form, such as mean number of sessions per week. It also would be desirable to contact subjects who did not posttest to find out what fraction were continuing to practice the various procedures learned.

We recommend that more emphasis be put on performance measures, which may be more reliable (or at least easier to evaluate) than self-report or physiological ones. There is also a need for more studies with long-term follow-up. If accurate estimates could be made of long-term attrition rates, cross-sectional comparisons of long-term practitioners might become more meaningful. We also strongly recommend that researchers not treat meditation as a generic independent variable, but should attend to and report the specific kinds of meditation that they are using.

Independently of these methodological considerations, the findings of the present study give grounds for optimism that at least some current treatment procedures can effectively reduce trait anxiety.

REFERENCES

ABRAMS, A. (1977). The effects of meditation on elementary school students. (Doctoral dissertation, University of California, Berkeley, 1976). *Dissertation Abstracts International, 37(9-A)*, 5689.

BASSETT, J., BLANCHARD, E., & ESTES, L. (1977). Effects of instructional-expectancy sets on relaxation training with prisoners. *Journal of Community Psychology, 5*, 166-170.

BRADLEY, B., & MCCANNE, T. (1981). Autonomic responses to stress: The effects of progressive relaxation, the relaxation response, and expectancy of relief. *Biofeedback and Self-Regulation, 6*, 235-251.

DILLBECK, M., & ORME-JOHNSON, D. (1987). Physiological differences between Transcendental Meditation and rest. *American Psychologist, 42*, 879-881.

FERGUSON, P. (1981). An integrative meta-analysis of psychological studies investigating the treatment outcomes of meditation studies. (Doctoral dissertation, University of Colorado, 1981.) *Dissertation Abstracts International, 42(4-A)*, 1547.

GLASS, G. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education, 5*, 351-379.

HEDGES, L. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics, 9*, 61-85.

LANE, D. & DUNLAP, W. (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology, 31*, 107-112.

LUPARELLO, T., LEIST, N., LOURIE, C., & SWEET, P. (1970). The interaction of psychologic stimuli and pharmacologic agents on airway reactivity in asthmatic subjects. *Psychosomatic Medicine, 32*, 509-513.

MOLTZ, D. (1977). Effects of internally generated sounds on mood. (Senior thesis, Princeton University, 1976.) Cited in P. Carrington, *Freedom in meditation*. New York: Anchor Press/Doubleday.

MORSE, D., MARTIN, J., FURST, M., & DUBIN, L. (1977). A physiological and subjective evaluation of meditation, hypnosis, and relaxation. *Psychosomatic Medicine, 39*, 304-324.

MORSE, D., MARTIN, J., FURST, M., & DUBIN, L. (1979). A physiological and subjective evaluation of neutral and emotionally-charged words for meditation (parts I-III). *Journal of the American Society of Psychosomatic Dentistry and Medicine, 26*, 31-38, 56-62, 106-112.

PAUL, G., & TRIMBLE, R. (1970). Recorded vs. "live" relaxation training and hypnotic suggestion: Comparative effectiveness for reducing physiological activity and inhibiting stress response. *Behavior Therapy, 1*, 285-302.

RIDDLE, A. (1979). Effects of selected elements of meditation on self-actualization, locus of control, and trait anxiety. (Doctoral dissertation, University of South Carolina, 1979.) *Dissertation Abstracts International, 40(7-B)*, 3419.

SARASON, I., & SPIELBERGER, C. (Eds.). (1975). *Stress and anxiety* (Vols. 1-4). Washington: Hemisphere.

SMITH, J. (1976). Psychotherapeutic effects of Transcendental Meditation with controls for expectation of relief and daily sitting. *Journal of Consulting and Clinical Psychology, 44*, 630–637.

SMITH, M., & GLASS, G. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752-760.

SPIELBERGER, C., GORSUCH, R., & LUSHENE, R. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

THROLL, D. (1981). Transcendental Meditation and progressive relaxation: Their psychological effects. *Journal of Clinical Psychology, 37*, 776-781.

ZUROFF, D., & SCHWARZ, J. (1978). Effects of Transcendental Meditation and muscle relaxation on trait anxiety, maladjustment, locus of control, and drug use. *Journal of Consulting and Clinical Psychology, 46*, 264-271.