Communication Education
Vol. 56, No. 1, January 2007, pp. 72–88

Routledge
Taylor & Francis Group

Testing Public Anxiety Treatments Against a Credible Placebo Control

Desiree C. Duff, Timothy R. Levine, Michael J. Beatty, Jessica Woolbright & Hee Sun Park

Research investigating public speaking anxiety treatments is subject to demand effects. This study tests the relative effectiveness of systematic desensitization (SD) and multiple treatment method (MT) containing visualization therapy against no-treatment and credible placebo controls. Data (n = 238) were collected at six points in a public speaking class. Reported anxiety decreased over time, but neither treatment outperformed the placebo control. Short-term gains were apparent for the MT group with participants selected on the basis of extreme high scores, but these gains did not persist relative to the control conditions over time. The current data suggest no evidence of lasting improvement of SD or MT treatments over a credible placebo control and enrollment in a standard public speaking class.

Keywords: Communication Apprehension; Systematic Desensitization; Placebo; Public Speaking

In *De Oratore*, Cicero (1942) confessed: "I turn pale at the outset of a speech and quake in every limb and in my soul" (p. xxvi). Scholars have been interested in the problem of public speaking anxiety ever since. Not surprisingly, much of the scholarship regarding public speaking anxiety has focused on methods for reducing it. At first glance, the literature appears to suggest that several approaches alleviate anxiety associated with public speaking (Daly, McCroskey, Ayres, Hopf, & Ayres, 1997). Several studies indicate that participants *report* substantially lower levels of anxiety after treatment (e.g., Ayres et al., 1993; Ayres & Hopf, 1987). A review of the anxiety reduction research employing physiological measures, however, presents a

Desiree C. Duff (M.A., Purdue University, 1982) is a doctoral candidate, Department of Communication, Michigan State University and Assistant Professor, Cornerstone University; Timothy R. Levine (Ph.D., Michigan State University, 1992) is Professor of Communication, Michigan State University; Michael J. Beatty (Ph.D., Ohio State University, 1976) is Professor of Communication, University of Missouri-St. Louis; Jessica Woolbright received her B.A. from Cornerstone University, 2005; Hee Sun Park (Ph.D., University of California at Santa Barbara, 2003) is Assistant Professor of Communication, Michigan State University. Tim Levine can be contacted at levinet@msu.edu

ISSN 0363-4523 (print)/ISSN 1479-5795 (online)
DOI: 10.1080/03634520601016186

somewhat less optimistic picture of therapeutic effectiveness. In fact, a meta-analysis of public speaking anxiety reduction studies indicates that treatment has less impact on physiological measures (e.g., heart rate) than on self-reports of speakers' anxiety (Allen, 1989).

One interpretation of the mixed results for self-report and physiological studies is that the apparent effectiveness of therapy might be more artifactual than real. Indeed, research indicates that systematic desensitization, for example, is not more effective than placebo inductions (Kazdin & Wilcoxen, 1976) when the placebo treatment is as credible as systematic desensitization. Despite the importance of establishing the internal validity of treatment studies, the therapies advocated in the communication literature have not been compared to highly credible yet bogus (i.e., placebo) treatment. In light of the high degree of confidence in treatment expressed in the literature, its methodological examination is critical. Although all treatment approaches should be tested against placebo inductions, the present study focused on the relative effects of placebo, systematic desensitization, and a multiple treatment strategy.

Treatment Effectiveness

Systematic Desensitization and Placebo Effects

Systematic desensitization (SD) was designed by Wolpe (1961) to reduce anxiety and phobias through the process of reciprocal inhibition (i.e., a person cannot be anxious and relaxed at the same time). Underlying SD is the view that anxiety can be extinguished through classical counter-conditioning. Clients are trained to relax while visualizing threatening scenes. Through the repeating pairing of the threatening stimuli and the relaxation response, relaxation eventually replaces anxiety as the dominant response to the stimulus. Although Kazdin and Wilcoxon's (1976) review of the treatment literature indicated that systematic desensitization was not more effective than an equally credible placebo treatment, we examined SD in this study because it continues to be advocated as an effective treatment in the communication literature (e.g., Friedrich, Goss, Cunconan, & Lane, 1997).

Allen's (1989) meta-analysis of the treatment literature raised serious questions about the effectiveness of SD, which point toward experimental artifacts. Among other things, Allen found significant differences in the effectiveness of various treatments, including SD, as a function of the type of dependent measurement used to assess effectiveness. The largest effect for SD was found for studies relying on self-reports of public speaking anxiety ($r = .34$). In contrast, the average effect for studies using physiological indices such as heart rate was more modest ($r = .18$) and significantly less than for self-report studies. These results led Allen to conclude that either researchers relying on self-report measurement "overestimate the effectiveness of therapy" or that physiological measurement "underestimates the effectiveness of treatment" (p. 137). According to Allen, either the observed results for self-report studies were inflated by "demand" characteristics or the results for "physiological

measures may be ambiguous (and therefore observe smaller effects) because positive excitement as well as fear is measured" (p. 137).

Although communication apprehension (CA) researchers have long acknowledged that physiological arousal can be subjectively experienced as either pleasant or unpleasant (e.g., Beatty & Behnke, 1991; Behnke & Beatty, 1981), the argument that physiological arousal can indicate positive excitement is not applicable to therapeutic studies. As Beatty and McCroskey (in press) pointed out, speakers who find public speaking as exciting in a positive way typically do not volunteer nor are they selected for treatment. Samples in SD studies most often consist entirely of highly apprehensive people who experience arousal during performance as a negative sensation. Therefore, small reductions in physiological arousal as an outcome of SD mean that treatment induced only a small counter-conditioning effect.

Allen further suggested that the results of his meta-analysis did not rule out the possibility that although arousal was not substantially reduced, participants might have relabeled it as positive excitement (1989, p. 137). This is unlikely for two reasons. First, the change in self-report scores does not indicate a relabeling of the high levels of arousal commensurate with a conversion to positive excitement. Research shows that persons high in public speaking anxiety are highly aroused during performance (Beatty & Behnke, 1991; Behnke & Beatty, 1981). The small change in physiological arousal induced by SD indicates that highly apprehensive persons remained highly aroused by public speaking after treatment. Whether they interpreted the arousal as positive or negative, their self-reports should correspond in magnitude for the post-treatment arousal levels. In other words, persons who were apprehensive and highly aroused before treatment should be highly positive about public speaking after treatment if they relabeled the high arousal. However, as Beatty and McCroskey (in press) pointed out, the shifts in self-report anxiety scores are far less than required to indicate a cognitive shift from anxiety to positive excitement.

Second, the mechanism for SD is not relabeling arousal but arousal reduction. In fact, SD contains no technique or feature that would induce relabeling of arousal. Therefore, even if relabeling were to occur, the mechanism affecting changes in public speaking anxiety would have nothing to do with reciprocal inhibition by counter-conditioning.

If the small effect of SD on physiological measures cannot be explained as ambiguous measurement or as relabeling, then the possibility that self-report results were inflated by demand characteristics in treatment studies becomes more plausible. Beatty and McCroskey (in press) noted that many of the features of systematic desensitization programs are precisely the conditions that researchers might hypothesize placebo effects. The clients are fully aware of the purpose of SD, have a stake in the outcome, are selected for extreme scores, and invest considerable time into the sessions. Further, SD programs are administered under the auspices of universities, by university personnel, and SD is described in scientific sounding terms (e.g., counter-conditioning, progressive relaxation, anxiety hierarchies). These problems also exist within other treatment approaches, including visualization and skills-based approaches. Therefore, demand effects such as those existing in treatment

environments could plausibly influence clients' subjective post-therapy reports. The degree to which self-report data are affected by demand characteristics can be estimated by comparing supposed therapeutic effects to those induced by placebo conditions, and comparing both to no-treatment, no-placebo control.

Visualization Therapy and Placebo Effects

Visualization therapy (Ayres & Hopf, 1987, 1989) consists of training participants in positive imagery techniques. Although the precise mechanism is unclear, advocates attribute the effectiveness of the treatment to increases in positive thinking coupled with reductions in negative thinking (Ayres et al., 1993; Ayres & Hopf, 1989). Nevertheless, research investigating the effectiveness of visualization techniques report that it is highly effective in reducing public speaking anxiety (Ayres et al., 1993; Ayres & Hopf, 1987, 1989, 1990). Visualization therapy alone (Ayres & Hopf, 1987; 1989; 1990) and combined visualization with skills training and SD (Ayres et al., 1993) are consistently reported to outperform both no-treatment controls (Ayres et al., 1993; Ayres & Hopf, 1987, 1989, 1990) and placebo controls (Ayres et al., 1993; Ayres & Hopf, 1989) both in the short term (Ayres et al., 1993, Ayres & Hopf, 1987, 1989, 1990) and over time (Ayres et al., 1993; Ayres & Hopf, 1990). Visualization has been found to be more effective than SD in some studies (e.g., Ayres & Hopf, 1989) but not in others (Ayres & Hopf, 1987). The short-term effect sizes for visualization reported in these studies are considerably larger than the average effects reported in Allen's (1989) meta-analysis, ranging for 23 to 57% of the variance in the visualization studies compared to 12% in the meta-analysis. The long-term effects, although smaller, are still substantial, ranging from 12–13% of the variance. These visualization studies, however, only investigate participants selected on the basis of extreme scores defined as one standard deviation above the mean. Thus, regression to the mean is possible.

The substantive content of the placebo conditions make it unclear if a valid comparison of visualization to placebo controls has been made in previous studies. Ayres and Hopf (1989) used two placebo controls. One included a SD procedure and the other involved rational thinking procedures. Although both groups resulted in less change than visualization, both placebos reduced apprehension substantially more than the no-treatment control. Whether these effects reflect demand effects or true therapeutic impact is unclear. The placebos in Ayres et al. (1993) included watching a videotape involving skills training or a video unrelated to public speaking. In these instances, the placebo group did not differ from a no-treatment control. It is unclear how watching an unrelated video is different from a standard no-treatment control. Thus, some previous placebo conditions contained potentially viable treatments, whereas others may have lacked credibility.

The necessary elements of a placebo condition are that it is bogus meaning that it does not reduce anxiety except as a function of demand or expectancy effects, and that it is credible—meaning that clients believe that it could be effective (Kazdin & Wilcoxon, 1976). No evidence was presented in the visualization studies to indicate

that various placebos met both of these criteria. In the systematic desensitization versus placebo literature, credibility of placebo treatments was critical. As mentioned, Kazdin and Wilcoxon demonstrated that SD was not more effective than *equally credible* placebo inductions. Furthermore, studies appear to support the relative effectiveness of SD only when placebo inductions were less credible than SD and, as Kazdin and Wilcoxon pointed out, differences in effectiveness were proportional to the differences in credibility of SD and placebo protocols.

If the results of a particular study indicate that placebo is as or more effective than a highly touted therapy, the effectiveness of the placebo is evidence that the placebo was credible. However, the conclusion that therapy is superior to placebo requires evidence independent of self-report anxiety scores to demonstrate that the credibility of the placebo protocols was comparable to that of the therapy condition. Otherwise, the possibility exists that differences in "effectiveness" merely indicate that the therapy protocols represent a more convincing placebo. In Ayres and Hopf (1989), visualization was compared to systematic desensitization and a placebo condition for which the credibility was not checked. Moreover, in the nearly 20 years since visualization was introduced to our discipline, no study designed to assess the physiological effects of therapy has appeared in the research literature.

Overall, it is unclear whether visualization has been adequately compared to credible placebo inductions. Therefore, the possibility that change in self-report scores associated with visualization therapy represents placebo effects remains an empirical question. In the present study, systematic desensitization and visualization are compared to an intuitively plausible placebo and no-treatment control conditions. Further, unlike most studies in the literature, participants were not selected on the basis of extreme scores, participants reported anxiety after delivering actual speeches to live audiences, and scores on public speaking anxiety were tracked over time. This final feature was important because demand effects might be relatively transient compared to true therapeutic improvement.

Method

Participants

The participants were 238 undergraduate students enrolled in a basic public speaking course at a small Midwestern university. The course was part of the core university requirements and was taught in sections consisting of approximately 24 students each. During the fall 2003 semester of the study, the course consisted of 10 sections, taught by seven different instructors. The majority of the enrollees were college freshmen (84%), with the minority of enrollees being sophomores (13.1%), juniors, seniors, or other (3%). Slightly more than 62% were female. Their average age was approximately 18.5 years ($M = 18.6$, $SD = 1.74$). The majority of students (94.1%) identified themselves as white-Caucasian. Data collection was Internal Review Board (IRB) approved.

As part of the course, all students were required to attend at least one—and no more than two—outside sessions. These sessions consisted of the various experimental conditions for the study. Students, however, were not required to participate in the study, and all students signed informed consent forms expressing their willingness to have their responses used in the study. Whereas no student actively declined to allow the researchers to use their responses, not all students completed all six data collections. The reduction in numbers of participants in the study as the study progressed reflected students who withdrew from the course, stopped attending, or did not complete course or research requirements (e.g., delivery of a speech, attendance at an outside session, or completing all the measures). In all, 238 participated in the pre-test, 224 completed both the pre- and post-test, and 185 completed all six data collections. Those failing to complete the post-test were disproportionately assigned to the control group, $\chi^2\ (3) = 10.83$, $p < .02$. Those who did not complete the post-test ($M = 3.04$), however, were not significantly different on pre-test anxiety scores from those completing the post-test ($M = 3.28$); $t\ (236) = 1.09$, $p = .28$.

Data Collection and Instrumentation

Students participating in the study were asked to complete six surveys at various times during the 16-week semester. Each survey was identical, consisting of the six-item subscale relating to the public speaking context from the Personal Report of Communication Apprehension-24, (PRCA-24, see Levine & McCroskey, 1990 for validity information). The pre-test ($M = 3.26$, $SD = 0.80$, distribution approximated normality, $\alpha = .87$), given near the beginning of the semester, included demographic items in addition to the six items from the PRCA. The second survey ($\alpha = .86$) was given immediately following the delivery of each student's first in-class speech of the semester approximately three weeks after the pre-test. The third survey ($\alpha = .86$) was given approximately five weeks later immediately following each student's second in-class speech. The fourth survey ($\alpha = .86$) was given immediately after each student's third in-class speech. The fifth survey ($\alpha = .88$) was given immediately after each student's final in-class speech of the semester. The third, fourth, and fifth surveys were spaced approximately three weeks apart. Lastly, the post-test ($\alpha = .89$) was given on the day of each student's final examination for the course, one to two weeks after the last speech. The intra-class correlation for the post-test measure as a function of class section was not statistically significant and small, indicating that the assumption of independence of observation was met, $\rho = .02$, $F\ (9, 214) = 0.72$, $p = .69$. The means for all six surveys are reported in Table 1.

Design and Experimental Conditions

Students were randomly assigned to one of four experimental conditions.[1] Those conditions included a no-treatment control condition, a credible placebo condition, and two types of treatment conditions. At the university, all students received laptop

Table 1 Mean Anxiety Scores by Time and Condition

Condition	*n*	Pre	S1	S2	S3	S4	Post	*n*
All Participants	238	3.26	3.34	3.05	3.07	3.01	2.96	185
Control	83	3.29	3.28	3.00	2.99	2.93	2.93	55
Placebo	54	3.13	3.23	2.93	2.91	2.83	2.73	47
Ayres Tape	45	3.34	3.48	3.17	3.24	3.20	3.07	35
SD	56	3.34	3.46	3.14	3.21	3.20	3.08	48

Note: The first *n* reflects pre-test scores, and the second reflects those completing all six data collections. The tabled means are based on $n = 185$.

computers at the beginning of the freshman year. Students were asked to bring their laptops to treatment sessions where they were lent a pair of headphones in order to view/listen to the content of their assigned treatment. Course instructors were not informed about the nature of the treatments, the existence of a placebo control, or which students were assigned to which condition.

The control condition received no treatment until the end of the semester. Specifically, these students attended an outside session during the time period between the delivery of their final speech and the postest. At that time, students were randomly assigned to receive one of the two types of treatments used in the treatment conditions (described below). This procedure was adopted for ethical considerations so that no student would be deprived of a potentially effective treatment.

Students in the placebo condition attended a session early in the semester, typically during the time between their first and second speeches. In that session, they were asked to listen to a compact disk (CD) of new age/world music for 15 minutes.[2] Participants were handed the CD and told that the CD not only contained music, but also contained subliminal messages. To enhance credibility, they were told that individuals may have varying degrees of sensitivity (limens or thresholds) for subliminal messages and were instructed to listen to the first few seconds of the CD in order to determine if they could detect these messages. They were told that if they heard "rushing" or "whispering" sounds that they were to remove the CD and insert a second CD with a lower subliminal threshold. Although an exact count was not kept, it appeared to the experimenter that most students listened to the first few seconds of the first CD, removed it, and replaced it with the second CD.[3]

In reality, the first and second CDs contained no subliminal messages. In fact, they were identical except for the initial few seconds on the first CD in which were recorded three statements: "Breathe," "Breathe slowly," and "Breathe deeply." These statements were digitized such that they were "laid under" the music at a level low enough to sound like a subtle "rushing" sound, perhaps like a sound made by an electronic instrument. The actual words were indistinguishable.

As in the control condition, students in the placebo condition were asked to return for a second session late in the semester, during the time following their final speech (preceding the post-test). Students were randomly assigned to receive one of the two types of treatments used in the treatment conditions (described below).

Two treatment conditions were included in the design. Participants in the first treatment condition were randomly assigned to attend an out-of-class session at one of two points in time, during the period of time between delivery of their first and second speeches, or the period of time between delivery of their second and third speeches. At the session, participants were individually exposed to a DVD containing a reformatted videotape created by Ayres and Ayres (1990) using a multiple treatment method (MT) approach to reducing public speaking anxiety. The reformatted tape entitled, "Coping with the Fear of Public Speaking," exposed viewers to three types of treatment techniques including visualization therapy (i.e., cognitive modification), skills-based management techniques, and systematic desensitization (SD). This videotape has been reported to be effective in reducing public speaking anxiety (Ayres et al., 1993).

Like the multiple treatment condition, participants in the second treatment condition were randomly assigned to attend an out-of-class session at one of the two points in time noted above. These students were exposed to systematic desensitization (SD). The session consisted of two parts. First, participants were exposed to a relaxation audiotape, reformatted to CD, developed for SD treatment of communication apprehension.[4] Second, in accord with SD protocol, the experimenter reentered the room and led the students through the scripted hierarchical SD exercises, in which students are asked to imagine themselves performing communication acts involving increasing levels of anxiety while maintaining a state of relaxation.

Results

The data were initially screened for evidence of selection artifacts and consistency with statistical assumptions. A one-way ANOVA on pre-test scores by condition revealed no evidence of selection effects, $F\ (3,\ 234) = 1.50,\ p = .21,\ \eta^2 = .02$. The distributions of the dependent measures were assessed with frequency distributions and Q-Q plots and did not deviate from normality. No evidence of heterogeneity of variance or non-independence of observations was observed. Within cell standard deviations ranged from 0.67 to 0.89. The largest intra-class correlation attributable to section or instructor was .06 and not statistically significant.

Next, scores for the six surveys were assessed with repeated-measures ANOVA, and a general decrease in levels of public speaking anxiety was observed from pre-test ($M = 3.26,\ SD = .80$) through the post-test ($M = 2.96,\ SD = .82$), with a slight increase following the participants' first speech ($M = 3.35,\ SD = .81$). The decline was statistically significant, $F\ (5,\ 920) = 26.26,\ p < .001,\ \eta^2 = .12$, and predominantly linear, $F\ (1,\ 184) = 49.46,\ p < .001,\ \eta^2 = .10$, with the linear contrast accounting for 78% of the explained sums of squares. Descriptive statistics for the six tests are summarized in Table 1.

The effects of treatment condition were assessed in several ways. First, the data were analyzed with a two-way mixed ANOVA with treatment condition as a fixed

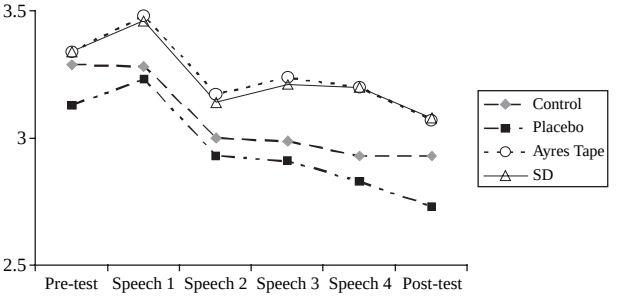

Figure 1. Mean Anxiety Scores for Experimental Groups over Time.

independent groups factor and time as a repeated factor. Neither the treatment condition main effect, $F\ (3,\ 181) = 1.87$, $p = .14$, $\eta^2 = .02$, nor the critical treatment by time interaction, $F\ (15,\ 905) = 0.56$, $p = .90$, $\eta^2 < .01$, was statistically significant. A plot of mean scores on apprehension over the course of the study, broken down by experimental group, is presented in Figure 1. Cell means are presented in Table 1.

Participants in the placebo, SD, and MT groups were directly compared on the test immediately following treatment and on scores on the second test (i.e., one lag) after treatment. No differences were observed between the three groups on the test immediately following the treatment, $F\ (2,\ 140) = 1.01$, $p = .37$, $\eta^2 = .01$, but a marginally significant difference was observed on the lagged test, $F\ (3,\ 135) = 3.07$, $p = .05$, $\eta^2 = .04$. Contrast analysis $(-2,\ +1,\ +1)$ suggested that scores in the placebo group were lower than the scores in the MT and SD groups, $t\ (135) = 2.44$, $p < .016$, $\eta^2 = .04$. Means are presented in Table 2.

A one-way ANOVA was used to examine scores at the time of the final speech. At this time, all participants in the two treatment conditions had completed treatment, and participants in the placebo and control conditions had not yet received MT or SD. The ANOVA failed to reveal significant differences at traditional levels, $F\ (3,\ 206) = 2.27$, $p = .08$, $\eta^2 = .03$. The order of means, from lowest to highest reported apprehension, was the placebo group ($M = 2.87$), followed by the control group ($M = 2.90$), MT group ($M = 3.16$), and the SD group ($M = 3.20$) respectively. A contrast analysis $(-1,\ -1,\ +1,\ +1)$ showed that together, participants in the placebo and control groups reported significantly lower anxiety than participants in the two treatment groups, $t\ (206) = 2.56$, $p = .001$, $\eta^2 = .03$. Together, these findings clearly show that SD and MT do not outperform a credible placebo control.

Highly Apprehensive Individuals

Most previous treatment studies select participants based on extreme scores (e.g., Ayres et al., 1993; Ayres & Hopf, 1987, 1989, 1992). While this practice raises concerns over regression to the mean, it could be argued that these are the individuals in greatest need of treatment and are those for whom the treatments were designed. To assess treatment effectiveness on the most apprehensive participants thereby

Table 2 Mean Anxiety Scores Following Treatment

Condition	Pre-test	Post-treatment	Lag
Placebo	3.13	2.94	2.91
Ayres Tape	3.24	3.16	3.24
SD	3.34	3.15	3.28

Note: Post-treatment scores are on the first test following treatment, and lag scores are scores on the next test. Discrepancies in means between tables are attributable to missing data.

directly replicating previous studies, the data were reanalyzed with only those individuals scoring one standard deviation or more above the mean on the pre-test. Thirty-five individuals met this criteria with scores greater than 4.00. Of these, 28 completed all measures. Due to the small sample size, these data should be interpreted with caution.

The high apprehensive data were initially analyzed with a two-way mixed ANOVA with treatment condition as a fixed independent groups factor and time as a repeated factor. As with the entire sample, the highly apprehensive individuals generally reported decreasing levels of anxiety over time, $F\ (5, 120) = 14.76$, $p < .001$, $\eta^2 = .16$, linear contrast, $F\ (1, 24) = 33.55$, $p < .001$, $\eta^2 = .15$. Neither the treatment condition main effect, $F\ (3, 24) = 1.96$, $p = .15$, $\eta^2 = .10$, nor the critical treatment by time interaction, $F\ (15, 120) = 1.10$, $p = .37$, $\eta^2 = .04$, was statistically significant. Cell means are presented in Table 3.

Participants in the placebo, SD, and MT groups were directly compared on the test immediately following treatment and on scores on the second test (i.e., one lag) after treatment. Statistically significant differences were observed between the three groups on the test immediately following the treatment, $F\ (2, 19) = 4.18$, $p = .033$, $\eta^2 = .33$. Contrast analysis $(+1, -2, +1)$ showed that scores in the MT group were substantially lower than the SD or placebo groups. These differences, however, were not evident on the lagged test, $F\ (2, 19) = 0.65$, $p = .53$, $\eta^2 = .07$, and a contrast analysis with the same weights was no longer significant, $t\ (135) = 1.14$, $p = .27$, $\eta^2 = .01$. Means are presented in Table 4.

Finally, a one-way ANOVA was examined for scores at the time of the final speech, and no statistically significant differences were observed, $F\ (3, 28) = 1.83$, $p = .16$, $\eta^2 = .16$. The order of means, from lowest to highest reported apprehension, was the MT group ($M = 3.41$), followed by the control group ($M = 3.47$), the SD

Table 3 Mean Anxiety Scores by Time and Condition for Highly Apprehensive Individuals

Condition	Pre	S1	S2	S3	S4	Post	*N*
All Participants	4.46	4.30	3.91	3.76	3.74	3.70	28
Control	4.48	4.26	3.67	3.57	3.43	3.40	9
Placebo	4.50	4.38	4.33	4.08	4.13	3.96	4
Ayres Tape	4.33	3.87	3.83	3.67	3.37	3.53	5
SD	4.48	4.52	4.00	3.98	4.07	3.93	10

Table 4 Mean Anxiety and Scores Following Treatment for Highly Apprehensive Individuals

Condition	Pre-test	Post-treatment	Lag
Placebo	4.50	4.33_a	4.08
Ayres Tape	4.33	3.19_a	3.73
SD	4.48	3.97	4.02

Note: Post-treatment scores are on the first test following treatment, and lag scores are scores on the next test.

group ($M = 4.00$), and the placebo group ($M = 4.12$) respectively. A contrast analysis ($-1, -1, +1, +1$) showed that together, participants in the MT and control groups reported significantly lower anxiety than participants in the placebo or SD groups, $t\ (28) = 2.26$, $p = .03$, $\eta^2 = .15$. These findings suggest that the MT procedure resulted in temporary reduction in communication apprehension for highly apprehensive participants, but that the improvements did not persist, nor were they superior to a no-treatment control.

Multilevel Analyses

Given that the research design involved repeated measures and that participants were nested within class sections and instructors creating possible non-independence of observation issues, the data were also examined with Hierarchical Linear Modeling were (Raudenbush & Bryk, 2002). These analyses also allow for partitioning of variance in CA scores into segments accounted for by stable individual differences, changes over time, and treatment conditions.

Initially, three-level nested structures were formulated. These consisted of multiple measures of CA for each individual (level 1, time-level) nested within individuals (level 2, individual-level) nested within sections or instructors (level 3, section-level or instructor-level). The intraclass correlation coefficients for the level-3 units were near zero. Class section accounted for only 0.012% of the total variance in CA scores (level-3 variance = 0.00008, $\chi^2[9] = 8.44$, $p > .50$), and the instructor accounted for only 0.12% of the total variance in CA scores (level-3 variance = 0.0008, $\chi^2[6] = 7.34$, $p = .29$).

Because the level-3 variance did not differ significantly from zero, subsequent analyses used a simplified two-level model. Level-1 (time, variance = 0.19877) accounted for 29.78% of the total variance in CA, and level-2 (individual differences, variance = 0.46861) accounted for 70.22% of the total variance in CA. Time was included as a linear predictor for CA, and was statistically significant, $b = -0.07$, $t = -7.54$, $p < .001$. Inclusion of time as a level-1 predictor reduced the level-1 variance from 0.19877 to 0.13753. Thus, 30.81% of the within-individual variance in CA was accounted for by the linear decrease in scores over time. Examination of the level-2 variance in the level-1 intercept (variance = 0.50032, $\chi^2[235] = 1825.89$, $p < .001$) and the level-2 variance in the level-1 slope for the time predictor (variance = 0.01212, $\chi^2[235] = 570.85$, $p < .001$) showed that there were significant

variations in the level-1 intercept (i.e., individuals had varying means on CA) and the level-1 slope (i.e., individuals had varying rate of change) across individuals. Thus, substantial individual differences were evident.

In the next step, treatments were introduced to see if differential assignment to treatment conditions explained either mean CA levels or the rate of CA reduction. The experimental conditions were dummy-coded. First, the control condition was coded as the reference group to contrast with each of the other conditions (i.e., placebo, the Ayres and Ayres tape, and SD). The results showed that none of the other conditions was significantly better in predicting the variations in the intercept or the slope than the control condition at $p < .05$. The second analysis included the placebo condition as the reference group. The results also showed that none of the other conditions was significantly better in predicting the variations in the intercept and the slope than the placebo condition at $p < .05$. The comparison of SD condition to the placebo condition, however, approached significance in predicting the intercept, $t\ (234) = 1.89$, $p = .06$, resulting in a 0.272 increase (standard error = 0.144) in individuals' average CA. For the third analysis, the control and placebo conditions were coded as the reference group, and the Ayres and Ayres tape and SD conditions were coded together as the comparison group. The results showed that the pooled treatment groups had a significant effect for the intercept, $t\ (236) = 2.12$, $p = .035$, resulting in a 0.209 increase (standard error = 0.099) in individuals' average CA. The comparison group, however, did not have a significant effect for the slope, $b = 0.015$, $t\ (236) = 0.79$, $p = .43$. Inclusion of this predictor reduced the level-2 variance in the intercept from 0.50032 to 0.49193; that is, only 1.68% of the individual variance in the average CA score was accounted for by the contrast between the two treatment conditions and the two controls.

Discussion

The present study was undertaken to determine whether the effects of public speaking apprehension treatment programs are confounded by placebo effects. Although published studies indicate that techniques such as systematic desensitization and visualization substantially reduce self-reported communication apprehension levels (e.g., Ayres et al., 1993; Ayres & Hopf, 1987), the possible contribution of placebo effects associated with the delivery of therapy have not been adequately examined. Studies in the psychological literature, however, indicate that systematic desensitization does not reduce anxiety more than a placebo treatment when SD and the placebo are seen by clients as equally credible. In the present study, participants in the placebo condition were told that subliminal anxiety-reducing messages, presented at precisely the correct limens, were embedded in music. In spite of the fact that the placebo treatments contained no such messages, the placebo control performed at least as well as the SD and MT treatments in reducing communication apprehension.

These findings are consistent with the research literature pertaining to the internal validity of treatment studies in general. In addition to replicating the findings reported in the placebo studies in the systematic desensitization literature (Kazdin &

Wilcoxon, 1976), our findings are consistent with Allen's (1989) meta-analysis. As mentioned earlier, Allen observed negligible reductions in physiological measures of anxiety for treatments. Certainly, physiological reactions are less susceptible to demand pressures than are self-report measures.

Allen speculated that one explanation for the small changes in physiological measures is that physiological responses such as heart rate can be ambiguous because arousal can indicate either anxiety or a positive affect such as excitement. However, Beatty and McCroskey (in press) noted that the magnitude of change reported is not consistent with high levels of arousal as indicative of positive affect. Beatty and McCroskey's point was observed by examining a scatterplot of pre- and post-test scores in the current data. Highly apprehensive individuals do not report a dramatic shift toward the other end of the scale, and the correlation between pre- and post-measures was $r\ (222) = .63$. Further, the multilevel analysis showed that the majority of the explained variance in communication apprehension scores over time was attributable to a stable individual difference. An alternative, as Allen noted, was that the relatively larger effects observed in self-reports indicates the presence of demand effects in treatment studies. Our findings are more consistent with the demand effect explanation.

Most research findings regarding the effectiveness of public speaking anxiety therapies are derived from assessments of ongoing programs rather than from experiments specifically designed to control for threats to internal validity. As a rule, such opportunistic strategies compromise the degree to which such threats can be controlled (Campbell & Stanley, 1963). On the other hand, our experiment exposed participants to less therapy than is typical of ongoing treatment programs. It would be tempting to diminish the importance of the results observed in the present study by supposing that larger effects for the therapies would have occurred given extended exposure to the various treatments. In addition to our results being consistent with previous research, it is worth noting the effect produced by our bogus music condition was not dissimilar to the average effect of full-term systematic desensitization programs. For example, Allen (1989) reported an average effect size (expressed as r), of .34. Thus, it is unlikely that extending the treatment program would produce effects for extant therapies that are substantially superior to placebo treatment, especially if the placebo treatments are also extended.

Our placebo treatment produced a change of 0.3 scale points in PRCA-24 scores. In previous research, Beatty, Behnke, and McCallum (1978) induced an *increase* of similar size in PRCA scores by merely informing participants that an impromptu speaking task would follow completion of the measure. It is rudimentary that fluctuations of this magnitude can result from the interaction of selection bias and demand characteristics. As already noted, samples for treatment studies in communication, for example, have been drawn exclusively from high scorers on measures such as the PRCA. Although the PRCA is psychometrically sound (Levine & McCroskey, 1990), the measure contains some error and, therefore, yields imperfect estimates of participants' true scores. When administered in a demand-free context, the error is assumed to be random. In pre-test/post-test designs involving

participants selected for their high pre-test scores, random response error results in regression toward the mean. Some researchers (e.g., Ayres, Hopf & Will, 2000) have argued that random assignment to control and treatment groups ensures that all groups regress toward the mean to the same degree. However, the argument is flawed in a fundamental way. Demand pressures, such as those induced during the delivery of treatment protocols, interact with selection bias effects producing larger effects in treatment than control conditions.

To illustrate the point, consider that during pre-testing, random response error manifests in high scorers through decisions regarding whether to select *strongly agree* or *agree* (or *strongly disagree* or *disagree*) in response to items. Highly apprehensive speakers are generally certain that they agree to some degree with the items but they are less clear whether *strongly agree* or *agree* is most accurate, at least for some items. In these types of scenarios, the choice between the two response options is random. However, pressures arising from demand properties of a treatment or experimental condition influence participants' choices in a particular direction. Within the context of therapeutic studies, demand characteristics systematically press participants to select the response indicative of less anxiety. Thus, treatment groups impose an additional artifactual influence above that normally expected due to regression effects. If this accentuated effect results from valid experimental inducement, it is usually referred to as "Selection × Treatment" interaction and is classified as a threat to external validity (Campbell & Stanley, 1963). However, if the catalyst is artifactual, such as a placebo effect or other confound, the phenomenon remains a threat to internal validity. Subtracting the amount of regression to the mean in control groups from that change in treatment groups when participants were selected for high scores does not provide protection against threats to validity. Rather, the practice simply amounts to subtracting the consequences of random response error from those of systematic error. The appropriate strategy under such conditions is to compare treatments to placebo conditions (Orne, 1962). If placebo conditions produce effects equal to or greater than extant therapies, then it is likely that the measured effects of therapies are due to the placebo qualities of the protocols used to deliver the therapies.

Generally, participants reported a statistically significant and linear decline in self-reported public speaking anxiety over the course of the study. Because all participants were enrolled in a standard public speaking course, the results indicate that the SD and MT did not produce improvement above that gained by the completion of a public speaking class alone. Because completing a public speaking class may have real therapeutic benefit resulting from practice and skills training, the current study lacked a true no-treatment control. Thus, it is possible that the treatments might lead to improvement relative to no intervention whatsoever.

Because all participants completed a standard public speaking course, it is tempting to attribute the general improvement to the efficacy of the class or skills training in general. Whereas the data are consistent with the explanation that completing a public speaking class systematically reduces public speaking anxiety, other explanations can plausibly account for this finding. It is possible that highly

apprehensive students were more likely to drop the class, and therefore the reduction in levels over time was attributable to differential mortality. Fortunately, this explanation, although plausible, can be ruled out. Pre-test anxiety scores for those students who completed the post-test ($M = 3.28$) did not differ from those who failed to complete the post-test ($M = 3.04$), $t\ (236) = -1.09$, $p = .28$, and the means were in the direction opposite to that predicted by the differential mortality explanation.

A third alternative explanation, however, is more likely. Given that all speeches were given in intact and relatively small classes, participants were likely to become increasingly familiar with the audience members over the semester. Rather than a reduction in trait anxiety, systematic reductions in reported anxiety might have been a function of a less threatening and more comfortable situation over time. The multilevel results provide evidence consistent with this alternative. Increased familiarity with the audience explanation is consistent with general decline across class sections and instructors.

It is interesting to note that the short-term effectiveness of the MT approach reported by Ayres and Hopf (1989) was replicated for subjects selected on extreme scores, but the relative effectiveness of this treatment did not persist over time relative to the no-treatment control. To the extent that the short-term effects for MT reflect experimenter demand or subject expectancy artifacts rather than true and lasting therapeutic improvement, the data are suggestive of an interesting treatment by subject interaction where students generally found the placebo credible, but highly apprehensive individuals found the MT approach more convincing. That is, people with different levels of trait anxiety may be differentially susceptible to different types of placebos.

Three limitations in the research merit specific comment. First, the credibility of the placebo was not formally checked. However, because the placebo performed as well or better than the MT and SD treatments, the placebo was presumably perceived as credible. Informal observations of the experimenter were also consistent with the contention that participants found the placebo highly credible. Second, only one application of SD was administered, and four or five applications are typical (Friedrich et al., 1997). It is possible that more sessions would produce better outcomes. Finally, and most important of all, the study did not include a long-term follow-up with a different audience. Such a follow-up would be necessary to determine if the general decline observed was attributable to lasting therapeutic improvement or mere familiarity with a specific context and audience.

In conclusion, this article tested the effectiveness of two treatments of public speaking anxiety. Scores on communication apprehension generally decreased over the course of the experiment, but it unclear if this decline was a function of effective skills training, actual practice, increased familiarity with the audience, or some combination of these elements. What is clear is that neither systematic desensitization nor visualization outperformed a credible placebo and a no-treatment control.

Notes

[1] The distribution of conditions across course sections and instructors were checked for uneven distributions, and the distributions did not depart from chance for either sections, χ^2 (27, $N=238$) $=20.75$, $p=.80$ or instructors, χ^2 (18, $N=238$) $=9.16$, $p=.96$. Further, all post treatment scores were checked for section or instructor effects. No significant intra-class correlations were found, and the largest was $\rho=.06$.

[2] Thom Jayne granted permission for the use of his music. To hear a sample of his music, one can visit the website at http://cdbaby.com/cd/thomjayne

[3] Although treatment credibility was not scaled (to avoid further demand effects), informal observations suggest that students found it highly credible. A number of students told the experimenter that they found procedures believable, and others expressed concerns about possible side effects.

[4] The tape was procured from Dr. James C. McCroskey, Department of Communication Studies, University of Alabama at Birmingham. It is currently available through the National Communication Association.

References

Allen, M. (1989). A comparison of self-report, observer, and physiological assessments of public speaking anxiety reduction techniques using meta-analysis. *Communication Studies*, *40*, 127–139.

Ayres, F. (Producer), & Ayres, J. (Writer). (1990). *Coping with the fear of public speaking* [videorecording]. Pullman, WA: Communication Videos.

Ayres, J., Ayres, F. E., Baker, A. L., Colby, N., De Blasi, C., Dimke, D., et al. (1993). Two empirical tests of a videotape designed to reduce public speaking anxiety. *Journal of Applied Communication Research*, *21*, 132–147.

Ayres, J., & Hopf, T. S. (1987). Visualization, systematic desensitization and rational emotive therapy: A comparative evaluation. *Communication Education*, *36*, 236–240.

Ayres, J., & Hopf, T. S. (1989). Visualization: Is it more than extra-attention? *Communication Education*, *38*, 1–5.

Ayres, J., & Hopf, T. S. (1990). The long-term effect of visualization in the classroom: A brief research report. *Communication Education*, *39*, 75–78.

Ayres, J., & Hopf, T. (1992). Coping with public speaking anxiety: An examination of various combinations of systematic desensitization, skills training, and visualization. *Journal of Applied Communication Research*, *20*, 183–198.

Ayres, J., Hopf, T., & Will, A. (2000). Are reductions in CA an experimental artifact? A Solomon Four-Group answer. *Communication Quarterly*, *48*, 19–26.

Beatty, M. J., & Behnke, R. R. (1991). Effects of public speaking trait anxiety and intensity of speaking task on heart rate during performance. *Human Communication Research*, *18*, 147–176.

Beatty, M. J., Behnke, R. R., & McCallum, K. (1978). Situational determinants of communication apprehension. *Communication Monographs*, *45*, 187–191.

Beatty, M. J., & McCroskey, J. C. (in press). The communibiological perspective. In Ayres, J., Hopf, T., Daly, J. A., McCroskey, J. C., & Ayres, D. M. (Eds), *Avoiding communication: Shyness, reticence, and communication apprehension* (3rd ed.). Cresskill, NJ: Hampton Press.

Behnke, R. R., & Beatty, M. J. (1981). A cognitive-physiological model of speech anxiety. *Communication Monographs*, *48*, 158–163.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Cicero, M. T. (1942). *De Oratore*. (E. W. Sutton, Trans.). Cambridge, MA: Harvard University Press.

Daly, J. A., McCroskey, J. C., Ayres, J., Hopf, T., & Ayres, D. M. (Eds.). (1997). *Avoiding communication: Shyness, reticence, and communication apprehension* (2nd ed.). Cresskill, NJ: Hampton Press.

Friedrich, G., Goss, B., Cunconan, T., & Lane, D. (1997). Systematic desensitization. In J. A. Daly, J. C. McCroskey, J. Ayres, T. Hopf, & D. M. Ayres (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (2nd ed., pp. 305–330). Cresskill, NJ: Hampton Press.

Kazdin, A. E., & Wilcoxon, L. A. (1976). Systematic desensitization and nonspecific treatment effects: A methodological evaluation. *Psychological Bulletin*, *83*, 729–759.

Levine, T. R., & McCroskey, J. C. (1990). Measuring trait communication apprehension: A test of rival measurement models of the PRCA-24. *Communication Monographs*, *57*, 62–72.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*, 776–783.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Wolpe, J. (1961). The systematic desensitization treatment of neuroses. *Journal of Nervous and Mental Disease*, *132*, 189–203.

Received May 5, 2006
Accepted September, 2006