A randomized trial of humor effects on test anxiety and test performance*

RONALD A. BERK and JOY NANDA

Abstract

This investigation tested the hypothesis of humor effects on test anxiety to improve test performance. A pretest-posttest control group design was employed to determine differences between humorous and serious versions of the same test content. One graduate biostatistics course of 98 students participated. Based on three independent test administrations, ANCOVAs were computed to isolate the effects of humorous directions only, humorous items only, and the combination of both on emotional/physiological and worry/cognitive anxiety symptoms and biostatistics achievement. Humorous directions had a statistically significant ($p < .05$) impact on constructed-response item performance for the first test (descriptive statistics), with an effect size of .43. Multiple-choice test performance correlated negatively with the two pre-anxiety subscales ($r = -.46$, $p < .001$), explaining up to 21% of the variance. The limitations of very low pre-anxiety levels and very high test performance precluded any other significant effects. The contributions of the humor technique used in the study and the value of measuring situation-specific anxiety immediately before and after a real testing condition were discussed.

Keywords: Humor; humor and test performance; humor in test items; test anxiety.

1. Introduction

Test anxiety correlates negatively with test performance in college courses (Cassady in press b; Hembree 1988; Musch and Broder 1999; Sarason

Humor 19–4 (2006), 425–450
DOI 10.1515/HUMOR.2006.[redacted]
0933–1719/06/0019–0425

1986). Students with high levels of test anxiety perform more poorly on all exams than their low-anxiety counterparts. Cognitive test anxiety accounts for approximately eight percent of the variance in student performance (Cassady and Johnson 2002). This variance is irrelevant to the construct being measured by a course test. Consequently, test anxiety represents a real threat to the validity of the achievement test scores used to assigned letter grades.

Test anxiety refers to transitory apprehensive, uneasy, or nervous feelings (affect state) immediately before, during, and after taking a specific test. For more than 30 years test anxiety research has concentrated on two dimensions: *emotionality*, which is manifested in the form of physiological symptoms, including rapid heart rate, nausea, dizziness, sweating, and fatigue, and *worry*, which refers to cognitive concerns about test taking and performance, such as negative expectations, preoccupation with performance, and potential consequences, which include the symptoms of self-critical, fear of failing, overwhelmed, and "going blank" (Deffenbacher 1980; Hembree 1988; Morris et al. 1981). The distinction between these dimensions and their impact on performance have been documented (Benson and Tippets 1990; Everson et al. 1991; Hong 2001; Hong and Karstensson 2002; Liebert and Morris 1967; Schwarzer 1984; Zeidner and Nevo 1992).

Recently, however, research on process models of test anxiety has challenged the preceding classic perspective. Such models broaden the conceptualization of test anxiety to include the patterns of behavior and cognitive responses during three phases in the learning-testing cycle: test preparation, test performance, and test reflection (Cassady in press a; Cassady and Johnson 2002; Schutz and Davis 2000; Schwarzer and Jerusalem 1992; Zeidner 1998). The second phase is the focus of this study.

Given what is currently known about test anxiety (Zeidner 1998), what steps can be taken to minimize its debilitating effects on performance? Typically, college or university counseling centers provide descriptions of the problem and list helpful tips, hints, and methods to reduce it. But, as educators, is there anything we can change in the tests themselves that can decrease test anxiety? In the domains of research concentrating on the psychological effects of humor (Berk 2001, 2002, 2004a, 2004b; Lefcourt 2001) and humor in course tests (Berk 2000, 2002; McMorris et al. 1997), there is mounting evidence of the potential positive effects of humor in the test directions and test items on reducing the symptoms of anxiety just prior to, during, and after test taking.

1.1. *Psychological effects of humor*

Among the numerous psychological effects of humor (Berk 2002), the one most pertinent to the conditions of testing in a college classroom is the reduction of the negative emotional consequences of anxiety (Abel 2002; Cann et al. 1999; Kuiper and Martin 1993, 1998; Kuiper et al. 2004; Nezlek and Derks 2001; Yovetich et al. 1990). Feelings of fear and worry take on new meaning when confronted with the testing experience. There is probably no other time throughout an entire semester when those negative emotions are at their peak as when the students walk into a class to take a test. Those emotions may even shoot off the chart when they see the test items.

The primary psychological function of humor is *detachment*. Psychological theorist Rollo May (1953: 61) stated that "Humor has the function of preserving the sense of self ... It is the healthy way of feeling a 'distance' between one's self and the problem, a way of standing off and looking at one's problem with perspective". The humor allows students to distance or detach themselves from the immediate threat — the TEST. The humor can reduce the negative feelings that would normally occur (Dixon 1980; Kuhlman 1984; O'Connell 1976). It also promotes a sense of objectivity and empowerment over the testing situation. In other words, humor can serve as an adaptive coping mechanism, and what better time for that mechanism to kick in, than during a close encounter of the testing kind? For a more detailed review of the psychological literature on humor, see Berk (2002) and Lefcourt (2001).

1.2. *Humor in course tests*

The empirical research on the effects of humor in testing has been critically reviewed by McMorris et al. (1997). They found only nine investigations of humor in college testing. All were conducted with students in undergraduate psychology classes. Seven used *content-irrelevant humor* in multiple-choice items (one of those also included short answer items), which is the most popular humor technique. It involves tacking humor onto a serious test by adding humorous distracters to several items or humorous items. The humor is irrelevant to the content of the test and the outcomes being measured. The other studies used anagrams with cartoons and humor in written dialogue between therapist and client.

McMorris et al. (1997) concluded that for the criterion of test performance, these studies provide insufficient and inconsistent evidence for using humor in tests to reduce anxiety and stress, and improve performance. Only the research by Smith et al. (1971) and Hedl et al. (1981) reported positive effects of humor on anxiety and stress reduction, respectively, and evidence of students' self-reported preferences for humor.

Considering the limitations of many of the investigations and the complexity of measuring interactions between humor in tests and other variables, McMorris et al. (1997) rendered the following verdict:

> Our own personal view at this juncture is to encourage the use of humor in tests, especially if instruction has included use of humor, the test has either no time limit or a very generous one, the humor is positive and constructive, the humor is appropriate for the group, test takers come from the same culture as the item writer, and the test developer feels comfortable in using humor. (McMorris et al. 1997: 295)

Since McMorris et al.'s (1997) review, there have been a few more studies. In one investigation by Perlini et al. (1999), humor frequency in the test items did not improve the test performance of highly test-anxious students. However, further analyses suggested that individual differences in the use of humor as a coping strategy significantly predicted exam scores. Another study by Bennett and Turner (2001) conducted over two and a half years found no significant effect from an additional humorous "E" alternative on multiple-choice test performance. Finally, in another multiyear study surveying students' self-perceptions about anxiety, 695 Johns Hopkins University undergraduate and graduate nursing students in seventeen biostatistics courses (six undergraduate, eleven graduate) over a six-year period indicated that humor was "very to extremely effective" in decreasing their test anxiety and improving their test performance (Berk 2000).

1.3. *Humor's link to test anxiety and performance*

What is humor's role in the context of the process models of test anxiety, mentioned previously?

1. There is evidence that students with high test anxiety are less likely to initiate effective coping strategies that could boost their test

performance levels (Onwuegbuzie and Daley 1996). Humor inserted into the test can serve as a coping mechanism by reducing the negative emotional and worry symptoms of test anxiety.

2. Students with high test anxiety, but good study skills, can experience the ''anxiety blockage phenomenon,'' where they report knowing the information before the test, but when they entered the room to take the test, the information mysteriously leaked out of their brains (Covington and Omelich 1987). The anxiety interferes with students' ability to retrieve information on demand once they open the test (Bar-Tal et al. 1999; Cassady in press a; Cassady and Johnson 2002; Mueller 1980; Naveh-Benjamin 1991). Humor on the cover of the test booklet or in the test itself may depress anxiety levels to reduce this blockage and retrieval processing failure that decrease performance.
3. High test anxious students encounter the perceived threat of the during the first few moments of testing. This may skew their judgments about the difficulty of the test, which can prompt self-deprecating thoughts, lack of concentration, and task-irrelevant thinking (Sarason 1986; Schutz and Davis 2000; Schwarzer and Jerusalem 1992). The test threat can also impair students' abilities to cope with the test experience, as noted above. The attendant anxiety levels can result in their failure to recall tasks and perform successfully on basic knowledge-level as well as higher-order reasoning test items (Cassady in press a). One of humor's primary psychological functions is to allow one to distance one's self from an immediate threat or aversive stimulus.

Humor's role in all of the above is to tackle test anxiety directly by effectively reducing its negative emotional consequences. But where does test performance fit into this hypothesis? The previous research suggests that humor affects performance indirectly by serving as a moderator variable. As humor decreases students' anxiety levels, their performance levels will increase. That is the framework within which this study examines humor effects on test anxiety and performance.

The aforementioned test anxiety studies have not measured the symptoms of anxiety immediately prior to and after testing. Can simply reading humorous test directions knock anxiety levels down a notch or two to eliminate retrieval blockage or ''going blank''? Can *content-relevant humor*, which is inserted into the actual content of different item formats, such as multiple-choice and constructed-response, decrease the symptoms of anxiety? It is now time to answer these questions and submit these humor effects to the rigor of a randomized trial.

The purpose of this study is to experimentally isolate the effects of humorous test directions and test items in reducing the anxiety students feel as they enter the testing environment and that occurring during the test itself. Can humor decrease test anxiety and, consequently, increase test performance?

1.4. *Hypotheses*

This study will test the following hypotheses:

1. Students exposed to humorous test directions will exhibit significantly lower anxiety and higher test performance than students exposed to serious directions.
2. Students exposed to content-relevant humor in the test items will exhibit significantly lower anxiety and higher test performance than students exposed to serious items.
3. Students exposed to humorous test directions and items will exhibit significantly lower anxiety and higher test performance than students exposed to serious directions and items.

2. Method

2.1. *Sample*

Students participating in this investigation were volunteers from a graduate biostatistics course at the Johns Hopkins University School of Nursing. It is a required course for master's degree and accelerated second-degree baccalaureate students, although seniors in the traditional baccalaureate program, master's and doctoral students from the school of public health, and special (non-degree) students also enroll. In total, 98 elected to take this course. The distribution by program was 31% traditional baccalaureate, 50% accelerated, 8% master's (nursing), 8% master's (public health), and 3% special. The mean age was 27 with 90% of the students female, 76% Caucasian, 10% African-American, 7% Asian, 4% Hispanic, and 1% Native American.

(AutoPDF V7 8/8/06 08:22) WDG (148×225mm) TimesM J-1582 Humor, 19:4 PMU: C(C) 31/07/2006 pp. 425–450 1582_19-4_03 (p. 430)

2.2. *Research design*

A pretest-posttest control group design was used to test the effects of humorous test directions, humorous items, and the combination of both on test anxiety and test performance. Consistent with routine test administration procedures over the past four years, all of the students in the course were randomly split into two groups and then randomly assigned to two lecture rooms for each of the three tests. The gender and ethnic distributions for the total class were characteristic of each randomized group $\pm$ two percent. Each room had a capacity of 110. The split allowed students more space "to spread out" and also minimized temptations to cheat.

The independent variables were types of test directions (humorous vs. serious), items (humorous vs. serious), and test (totally humorous [directions & items] vs. totally serious). For each independent variable, two new randomized groups were created with a new achievement test. Although the same 98 students participated in each phase, the design was structured as three separate studies. The dependent variables consisted of test anxiety, measured by the Symptoms of Test Anxiety Scale (STAS), which contained two subscales, and biostatistics achievement, measured by three different tests administered at three different time points during the course. The items were partitioned into multiple-choice and constructed-response sections. The covariates or premeasures consisted of algebra ability, measured by the Basic Algebra Proficiency (BAP) test, and the pretest administrations of the STAS.

2.3. *Instruments*

Ten different instruments were employed in this study. One provided baseline information on algebra ability. Three furnished baseline and outcome measures of test anxiety. The other six tools consisted of two versions of three different biostatistics tests. Each instrument is described below.

Basic Algebra Proficiency (BAP). This short 10-item test measures the specific algebra skills required in the computations of formulas used in the statistics course. It was used previously in a study of humorous teaching strategies in undergraduate and graduate biostatistics courses (Berk and Nanda 1998). Item difficulties ranged from 70–94%, item-total r ranged from .42–.66, and the coefficient alpha was .82 ($n = 152$). In this study, only one graduate biostatistics course with very high ability students was selected. Six of the 10 original items were chosen based on the item analysis: item difficulties ranged 89–98%, item total r ranged .31–.44, and coefficient alpha was .63 ($n = 89$).

Symptoms of Test Anxiety Scale (STAS). Available scales of test anxiety measure emotionality, worry, and, most recently, cognitive processes associated with test anxiety. They contain from 18 to 40 statements with dichotomous or polytomous response anchors and require 5 to 20 minutes to complete. The most frequently used scales are the Test Anxiety Scale (37 items) (Sarason 1980), Test Anxiety Inventory (20 items) (Spielberger 1980), Reactions to Tests Scale (40 items) (Sarason 1984), Revised Test Anxiety Scale (18 items) (Benson et al. 1992), and Cognitive Test Anxiety Scale (27 items) (Cassady and Johnson 2002). Unfortunately, the structure and length of these tools preclude their administration *immediately* prior to or after a test; they are usually given a couple of days before a test.

A new instrument was constructed for this study that could measure the physiological/emotional and psychological/worry/cognitive dimensions of test anxiety as students entered the classroom to take the test *and* be completed within one to two minutes prior to being handed the test. This Symptoms of Test Anxiety Scale (STAS) originally contained two 20-item lists of signs and symptoms of test anxiety drawn systematically from the aforementioned scales and the descriptions of state and trait test anxiety in the research (see Appendix A). The items consisted of one or two words in checklist format. Students were asked simply to check prior to the test "each characteristic related to taking this test that describes you right now before you begin the test" and after the test "that describes you right now after the test." A check represented the presence of a symptom and a blank, the absence, coded 1 or 0. This dichotomous response mode was used instead of polytomous anchors that measured the intensity of symptoms in order to facilitate ease and speed of response. List 1 contained the physiological/emotional (PHYS) symptoms; list 2 presented the psychological/cognitive (PSYCH) signs. The higher the score on each list, the higher the level of anxiety.

Item and reliability analyses of this 40-item, dichotomous-response scale were conducted for each test administration on pre-STAS data only. Different item statistics produced different PHYS and PSYCH subscales at each administration ($n = 97$):

Administration	Subscale	No. of Items	Item-Total r	Coefficient α
1	PHYS	6	.22–.45	.61
	PSYCH	10	.22–.55	.61
2	PHYS	10	.22–.42	.64
	PSYCH	8	.37–.69	.80
3	PHYS	10	.19–.44	.66
	PSYCH	11	.24–.59	.70

Across the three administrations or versions of each subscale, there were only two recurring PHYS symptoms (trembling hands and rapid heart rate) and four PSYCH symptoms (worried, "going blank," racing thoughts, and panicky). There were seven items on the PHYS and eight on the PSYCH subscale that appeared on two of the three versions. Three PHYS items and five PSYCH items occurred only once.

The validity coefficients between each subscale pair were .22, .47, and .66 for the three respective testings. The percentage of unexplained variance in each case (56–95%) suggested that the PHYS and PSYCH subscales were measuring different types of anxiety.

Biostatistics Achievement. There were three statistics tests in the course which were nonredundant in content coverage. They were weighted as 20%, 30%, and 50% for grading purposes. All tests were administered in open-book, open-everything format with calculators required. It has been found that students with high test anxiety are even outperformed on open-book take home examinations (Benjamin et al. 1981). These statistics tests were designed as power tests (rather than speed) so that all students could finish in the allotted time. No memorization, knowledge level items were included. The tests measured abilities to understand, apply, analyze, and make statistical decisions as a researcher. The items simulated *research problem-solving* with multiple-choice, matching, and constructed-response formats. This achievement outcome variable was measured with total score, multiple-choice (with matching) items subscore, and constructed-response items subscore. The percentage of constructed-response item points increased with the complexity of the statistics content on the three tests: *Test 1* (CR = 65%), *Test 2* (CR = 73%), and *Test 3* (CR = 80%).

(AutoPDF V7 8/8/06 08:22) WDG (148×225mm) TimesM J-1582 Humor, 19:4 PMU: C(C) 31/07/2006 pp. 425–450 1582_19-4_03 (p. 433)

Test 1, administered after about one month of classes, covered frequency distributions and graphs, levels of measurement, and measures of central tendency and dispersion. There were 57 total raw score points across all item formats: 20 for multiple-choice/matching and 37 for constructed-response. Two versions of this test were administered: *Test 1A*, which contained serious directions and items, and *Test 1B*, which had humorous directions on the cover, but serious items identical to those on *1A*. These humorous directions are shown below:

> GENERAL DIRECTIONS
>
> Sit down and make believe you're at the beach.
>
> Place the ANSWER SHEET somewhere in front of you but NOT in the sand. Print your name, social security number, current blood pressure and pulse rate, cholesterol level (HDL & LDL), triglicerides, and test booklet number in the upper right corner. Read the directions for marking your answers.
>
> Answer all questions as best you can. There will be no penalty for guessing, so guess away. You will have the entire class period to complete the test, which means you have 1.25 minutes per question. Pace yourself accordingly.
>
> DO NOT begin the test until you are told to do. I am going to let you sit here and sweat in the sun for about 30 minutes before letting you start the test. You are allowed to breathe; but nothing else. Watch out! Here comes a wave!

Test 2, administered a month later, measured *z*- and *T*-score transformations, Pearson correlation, six other types of correlation, simple linear regression, and multiple regression. This test totalled 60 points: 16 for multiple-choice/matching and 44 for constructed-response. *Test 2A* was the all-serious version; *Test 2B* had serious directions, but content-relevant humorous items throughout all sections. Content-relevant humor is an integral part of the item content, not an add-on (content-irrelevant), such as a choice E (see Berk 2000, 2002). Examples of humorous and nonhumorous items are given in Appendix B.

Test 3, administered during one three-hour block on the last class of the semester, assessed survey sampling statistics, power analysis, three *t*-tests, one-way analysis of variance and multiple comparison tests, and chi square and related nonparametric statistics. All statistics were taught and tested in the context of problem solving and complete experimental design structures. There were 100 total points: 20 for multiple-choice/matching and 80 for constructed-response. *Test 3A* was all serious; *Test 3B* contained humorous directions (different from *Test 1B*) and humorous content-relevant items (similar in form, but different in substance from those in *Test 2B*).

2.4. *Procedures*

A week before *Test 1*, the two randomized groups of students were announced in class and posted. The treatment-control conditions were randomly assigned to the two groups. Two TAs and one TA and the professor administered all instruments in the two respective rooms following these steps:

1. As students entered each classroom, a TA checked off the student's name on the roster to make sure he or she was in the correct room and then handed the student the pre-STAS. Completing the scale was optional. The student received a half a bonus point on the test if he or she consented to completing it. An additional half a point could also be earned by completing the scale again on the last page of the test.
2. As each student completed the scale and raised his or her hand, the other TA or professor collected the scale, handed the student a test booklet, and checked off the student's name on the class roster.
3. Once a student completed the test, he or she had the option of completing the scale on the last page of the test. When the test booklet was turned in to a TA, she verified the scale completion (or not) and checked off the student's name on the class roster.

This entire procedure was repeated for *Tests 2* and *3*. Even the randomization was repeated to minimize any testing effect or bias from the humor treatment and to maximize the differences and independence of the three treatments. Each test administration represented a separate study.

The administration schedule for all of the aforementioned tests and scales at key points during the biostatistics course is summarized below:

Course time Class 1	*Humor treatment* BAP test	*Serious control* BAP test
Test 1	Pre-STAS	Pre-STAS
(Descrip. Stat.)	*Humorous Directions*	Serious Directions
	Serious Items	Serious Items
	Post-STAS	Post-STAS
Test 2	Pre-STAS	Pre-STAS
(Corr/Regress)	Serious Directions	Serious Directions
	Humorous Items	Serious Items
	Post-STAS	Post-STAS
Test 3	Pre-STAS	Pre-STAS
(Exper. Design)	*Humorous Directions*	Serious Directions
	Humorous Items	Serious Items
	Post-STAS	Post-STAS

2.5. *Statistical analyses*

The pretest-posttest control group design was intended to isolate the effects of the three independent variables identified in Hypotheses 1.0, 2.0, and 3.0: (1) humorous vs. serious directions, (2) humorous vs. serious test items, and (3) humorous vs. serious directions and items. Basic Algebra Proficiency (BAP) and the pre-Symptoms of Test Anxiety Scale (pre-STAS) served as covariates to adjust posttest means for any initial between-group differences, thereby increasing the precision of the comparisons. The pre-STAS PHYS and PSYCH subscales were used as separate covariates in the analyses rather than being combined into one total scale. This decision was based on the low validity coefficients between the subscales, indicating that they may be measuring different types of anxiety and, thereby, produce different effects compared to their combination. The dependent variables were Biostatistics Achievement total, multiple-choice (MC) only, and constructed-response (CR) only on the three different tests and post-STAS.

The three hypotheses were tested with analyses of covariance (ANCOVA). Since each hypothesis related to a separate study, there were five ANCOVAs per study (two on anxiety and three on achievement). The pre-STAS PHYS subscale served as the covariate for the post-STAS PHYS subscale results in this pre-post design; the pre-STAS PSYCH subscale served a parallel function for the post-STAT PSYCH results. The remaining ANCOVAs tested the effectiveness of humor as a moderator variable to improve achievement test performance. Since previous research has indicated that test anxiety correlates negatively with test performance, if the humor can decrease the students' test anxiety, their performance should increase.

Levene's homogeneity of variance test was calculated to assess whether the homogeneity of variance assumption was violated. Statistical power for the F test for each analysis with an average $n = 49$ per group, $\alpha = .05$, and medium effect size = .50 standard deviation, was estimated to be 80%.

3. Results

3.1. *Preliminary analyses*

First, the option given to each student to participate in each phase of the study and receive a bonus point or not to participate resulted in 100% participation. The bonus point for completing the pre- and post-anxiety scale was insignificant out of the total number of points on each test and the students' very high levels of performance. No borderline grade was inflated to the next higher grade on any of the tests as a result of the bonus point. Consequently, that participation point had a negligible, if any, effect on the results.

Next, prior to computing the ANCOVAs to test the hypotheses, basic descriptive statistics were calculated for all variables in the design, plus Pearson correlation coefficients between the covariates and dependent variables. Tables 1 and 2 display the unadjusted pretest and posttest means and standard deviations and Levene's statistics for homogeneity of variance for the Symptoms of Test Anxiety subscales and Biostatistics Achievement (posttest only) for humorous and serious test samples.

Table 1 indicates that the students were far from petrified as they began and ended each test. Based on the number of symptoms (items) on each subscale (shown in parentheses), pretest *M* anxiety for both samples ranged from about .5–1.5 symptoms, but with relatively high standard deviations. PHYS and PSYCH anxiety was lowest entering *Test 2* (correlation/regression) and highest entering *Test 3* (final exam on experimental design). More than 50% of each group had 0 anxiety on both subscales prior to *Tests 1* and *2*. That percentage dipped to 39–42% before the final.

Table 1. Anxiety: *Unadjusted pre- and post-STASubscale means and standard deviations and Levene's homogeneity of variance test for humorous and serious test samples*

Dependent variable (items)	*Humorous test* ($n = 49$)[c]		*Serious test* ($n = 49$)[d]		Levene's HV (Pre)
	Pre-*M* (*SD*)	Post-*M* (*SD*)	Pre-*M* (*SD*)	Post *M* (*SD*)	
STAS 1					
a. *PHYS* (6)	.98 (1.35)	.50 (.89)	1.02 (1.25)	.37 (.57)	.19[1]
b. *PSYCH* (10)	1.08 (1.38)	.63 (.93)	.78 (1.30)	.39 (.64)	.16
STAS 2					
a. *PHYS* (10)	.67 (1.06)	.67 (.88)	.90 (1.43)	.69 (.95)	1.08
b. *PSYCH* (8)	.46 (1.18)	.63 (1.21)	.51 (1.28)	.46 (.82)	.11
STAS 3					
a. *PHYS* (10)	1.59 (1.91)	.94 (1.13)	1.02 (1.23)	.69 (.72)	8.08[b]
b. *PSYCH* (11)	1.51 (2.10)	.84 (1.26)	1.00 (1.20)	.75 (1.00)	5.90[a]

Key: [a]$p < .05$; [b]$p < .005$; [c]($n = 48$) for STAS 2 results; [d]($n = 48$) for STAS 3 results.

With the pretest anxiety levels so low, how much of a decrease could possibly occur? All posttest *M*s were less than one symptom and most, especially those after the first two test administrations, hovered above and below .5 symptoms. Although several pretest *M*s dropped by more than 50% or .5 symptoms on the posttest during *Tests 1* and *3*, these relative changes are negligible compared to the possible symptom score on each subscale. For the students in the two samples, pretest to posttest *M* anxiety decreased from low to really low after each treatment, despite the wide range of symptoms within each sample. The humorous test sample exhibited consistently higher variances than the serious test sample for *Tests 1* and *3*. Even the homogeneity of variance assumption on the presubscale score was violated prior to the final exam. The lowest pretest and posttest anxiety levels and variances were found for *Test 2*.

Table 2 shows the effects of negligible test anxiety — very high mean performance on every test with relatively low standard deviations. The homogeneity of variance assumption was violated twice for the first test only, on total score and constructed-response items.

Inasmuch as the unadjusted posttest between-sample mean differences appeared so small, how could the precision of the ANOVA comparisons be improved with covariates? The most appropriate choice for the post-STAS was pre-STAS. The correlations between pretest and posttest PHYS subscales for the three administrations ranged from .28–.51; pretest and posttest PSYCH subscales ranged from .39–.54. These coefficients were all statistically significant ($p < .05$) and computed on the

Table 2. Achievement: *Unadjusted post-biostatistics achievement means and standard deviations and Levene's homogeneity of variance test for humorous and serious test samples*

Dependent variable (items)	*Humorous test* *(n = 49)*[c]	*Serious test* *(n = 49)*[d]	Levene's HV
	M (*SD*)	*M* (*SD*)	
Biostat Achiev 1 (57) (Descript. Stat.)	54.34 (1.93)	53.90 (2.58)	7.12[a]
a. *Multi-Choice* (20)	18.16 (1.63)	18.45 (1.50)	.47
b. *Const-Resp* (37)	36.11 (1.03)	35.49 (1.90)	11.75[b]
Biostat Achiev 2 (60) (Corr/Regress)	57.71 (2.09)	57.20 (3.34)	2.60
a. *Multi-Choice* (16)	15.02 (1.18)	14.76 (1.82)	1.52
b. *Const-Resp* (44)	42.69 (1.57)	42.45 (2.60)	3.43
Biostat Achiev 3 (100) (Exper. Design)	94.90 (3.75)	94.00 (3.67)	.02
a. *Multi-Choice* (20)	17.88 (1.48)	17.35 (1.99)	3.17
b. *Const-Resp* (80)	77.02 (2.93)	76.67 (3.03)	.76

Key: [a]$p < .01$; [b]$p < .001$; [c]$(n = 48)$ for Biostat Achiev 2 results; [d]$(n = 48)$ for Biostat Achiev 3 results.

three control (serious test) groups only. Since the anxiety scale was administered immediately prior to and after the test, the intervening time to expect any changes from the humorous directions and/or items ranged from 1.5–2 hours for the first two tests and 2–3 hours for the third.

The covariates for the Biostatistics Achievement comparisons were supposed to be Basic Algebra Proficiency (BAP) and pre-STAS. The correlations between BAP and the three achievement tests were too low (and nonsignificant) to justify the use of algebra ability as a covariate. The *r*s ranged from .13–.25 for *Test 1*, 0–.21 for *Test 2*, and 0–−.08 for *Test 3*. The students scored so high on the test ($M = 5.51/6$) that there was minimal variance ($SD = .98$).

Instead, the covariates of choice were the pre-anxiety subscales. They produced higher and negative correlations with test performance. The highest significant correlations of PHYS and PSYCH with Biostatistics Achievement were with the multiple-choice items on every test, except one, the total score on *Test 1*. They ranged from −.19 to −.46. In other words, consistent with previous findings (e.g., Hong and Karstensson 2002), pre-anxiety symptoms on both PHYS (emotionality) and PSYCH (worry/cognitive) subscales correlated negatively with biostatistics test performance. They explained as much as 21% of the variance on the multiple-choice section of *Test 2* ($r = -.46$, $p < .001$). Consequently, the

(AutoPDF V7 8/8/06 08:22) WDG (148×225mm) TimesM J-1582 Humor, 19:4 PMU: C(C) 31/07/2006 pp. 425–450 1582_19-4_03 (p. 439)

Table 3. Anxiety: *Analyses of covariance of humorous vs. serious directions/items on three post-STASubscales*

Dependent variable	Source of variation	*SS*	*df*	*MS*	*F*
Post-STAS 1					
a. *PHYS*	Hum Direct	.41	1	.41	.80
	Error	48.75	95	.51	
b. *PSYCH*	Hum Direct	.65	1	.65	1.26
	Error	48.87	95	.51	
Post-STAS 2					
a. *PHYS*	Hum Items	.58	1	.58	.93
	Error	58.09	93	.63	
b. *PSYCH*	Hum Items	.86	1	.86	.98
	Error	81.80	93	.88	
Post-STAS 3					
a. *PHYS*	Hum Dir + It	.13	1	.13	.20
	Error	61.67	94	.66	
b. *PSYCH*	Hum Dir + It	.15	1	.15	.19
	Error	75.01	94	.80	

two anxiety subscales were used as the covariates for each hypothesis test where biostatistics achievement was the dependent variable.

3.2. *Hypothesis 1.0 (humorous directions)*

The effect of humorous vs. serious test directions on reducing pre- to posttest PHYS and PSYCH anxiety on the STAS was measured by the first two ANCOVAs in Table 3. Neither F ratios reached significance. The humorous directions didn't have any more impact on decreasing the students' anxiety during the test than the serious directions. The null hypothesis of equal adjusted *M*s on the STAS subscales could not be rejected. This part of the research hypothesis was not supported.

The humorous directions, however, did affect test performance. The results of three ANCOVAs on the first Biostatistics Achievement test on descriptive statistics are shown in Table 4. Among the three analyses, the humorous directions produced a significant difference ($p < .05$) on the constructed-response section of the test. Although the performances by both samples were extremely high, the adjusted means ($M_H = 36.13$, $M_S = 35.47$) yielded a difference of .66 and an effect size of .43 standard deviations. The null hypothesis for the constructed-response items was rejected, but not for the multiple-choice items and total test. Consequently, the research hypothesis was partially supported.

(AutoPDF V7 8/8/06 08:22) WDG (148×225mm) TimesM J-1582 Humor, 19:4 PMU: C(C) 31/07/2006 pp. 425–450 1582_19-4_03 (p. 440)

Table 4. Achievement: *Analyses of covariance of humorous vs. serious directions/items on three biostatistics achievement tests*

Dependent variable	Source of variation	*SS*	*df*	*MS*	*F*
Biostat Achiev 1	Hum Direct	7.78	1	7.78	1.62
(Descript. Stat.)	Error	451.78	94	4.81	
a. *Multi-Choice*	Hum Direct	.94	1	.94	.40
	Error	218.78	94	2.33	
b. *Const-Resp*	Hum Direct	10.57	1	10.57	4.66[a]
	Error	213.38	94	2.27	
Biostat Achiev 2	Hum Items	2.68	1	2.68	.40
(Corr/Regress)	Error	630.36	93	6.78	
a. *Multi-Choice*	Hum Items	.62	1	.62	.36
	Error	160.18	93	1.72	
b. *Const-Resp*	Hum Items	.68	1	.68	.15
	Error	427.12	93	4.59	
Biostat Achiev 3	Hum Dir + It	20.52	1	20.52	1.48
(Exper. Design)	Error	1288.67	93	13.86	
a. *Multi-Choice*	Hum Dir + It	9.54	1	9.54	3.15
	Error	281.78	93	3.03	
b. *Const-Resp*	Hum Dir + It	1.81	1	1.81	.20
	Error	829.72	93	8.92	

Key: [a] $p < .05$.

3.3. *Hypothesis 2.0 (humorous items)*

The ANCOVA results for this hypothesis of differences in anxiety reduction on both the PHYS and PSYCH subscales in Table 3 produced no significant F ratios. The humorous multiple-choice and constructed-response items had no differential impact on test anxiety. The level of posttest anxiety was slightly more than .5 symptoms for three of the means, and less than .5 for the other. The second test produced the lowest pre-anxiety levels and smallest changes among the three testing conditions. Based on these results, the null hypothesis could not be rejected.

Similar results were found for the correlation and regression test. The three ANCOVAs in the middle of Table 4 indicated that the humorous items did not significantly improve test performance compared to the serious items. The mean performances were extremely high in both groups and almost identical. These findings again led to nonrejection of the null hypothesis and no support for the research hypothesis.

(AutoPDF V7 8/8/06 08:22) WDG (148×225mm) TimesM J-1582 Humor, 19:4 PMU: C(C) 31/07/2006 pp. 425–450 1582_19-4_03 (p. 441)

3.4. *Hypothesis 3.0 (humorous directions and items)*

Given the trend in the preceding results, the combination of humorous directions and items did not significantly reduce anxiety or increase test performance more than the totally serious test. The ANCOVAs at the bottom of Tables 3 and 4 yielded nonsignificant F ratios. The only interesting exception on this experimental design final exam was the multiple-choice section. The performance was so high and the within-group variances were so low that the F ratio between the two groups was nearly significant with an effect size of .36. Despite this near hit, the nulls for the anxiety and achievement test dependent variables were not rejected and the research hypothesis was not supported.

4. Discussion

4.1. *Where's the test anxiety?*

What happened? What was I thinking? How can you study test anxiety when it doesn't occur? Could you find a more conservative real-world course condition to study test anxiety? Probably not, because there were specific systematic strategies used in the course to reduce test anxiety and maximize test performance and success. Consider the following:

1. Design "power" tests with adequate time limits so students don't feel rushed and everyone finishes.
2. Design open-book, notes, everything tests that measure higher-order thinking skills.
3. Design tests with a variety of item formats, if feasible, so students who have difficulty with multiple-choice will have other options.
4. Provide adequate test review information, a formal in-class or out-of-class review, and/or pep rally to pump students up before the test (see Berk 2002).
5. Have students pick the test date, if feasible, by majority vote so they have few or no competing tests or projects due for other courses (see Berk 2002).
6. You and/or your TAs should be available the week before the test.

All of these techniques were implemented during the course being studied. Evidently, they were more effective than the humor in the tests.

The pre-anxiety levels were so low before each test that there was little need for any intervention to reduce the levels further.

4.2. *How could test performance improve?*

As if nearly test anxiety-free students wasn't bad enough for this study, their performance on all three tests didn't help much either. Both samples aced every test with very low within-sample variance. If humor is hypothesized as a moderator variable to improve test performance, there needs to be some room for improvement. On this second count, the ceiling effect and restriction in range at the upper end of each score scale provided minimal wiggle room for differences in performance between samples.

4.3. *Humor effects*

Despite extremely low pretest and posttest anxiety and extremely high test performance, what new information does this study contribute? The only significant finding was the effect of humorous directions on test performance, specifically the constructed-response items on the first test on descriptive statistics. And the effect size was a nontrivial .43, almost a half a standard deviation. Since no previous research has studied humor in directions or constructed-response tests, this result provides preliminary evidence of the potential effect humorous test directions can have on test performance.

This humor effect is particularly noteworthy because there was no significant decrease in anxiety, which suggests that the humor was not serving as a moderating variable. Despite the major limitation of test anxiety-free students as they began the test, the humorous directions may have spiked their level of attention, interest, alertness, memory, or overall mental functioning as they began answering the questions, which produced an improvement in performance. Reading something funny just before taking the test may have less of a direct impact on test anxiety and more of an effect on mental processing. Since the same basic cognitive process is involved in the resolution of incongruity humor and problem solving (Goldstein et al. 1975; Johnson 1990; Suls 1972, 1983) and 65% of the test consisted of constructed-response statistics problems, the humorous

(AutoPDF V7 8/8/06 08:22) WDG (148×225mm) TimesM J-1582 Humor, 19:4 PMU: C(C) 31/07/2006 pp. 425–450 1582_19-4_03 (p. 443)

directions may have primed or jump-started the students' right hemispheres, which translated into improved problem-solving performances on the test. This effect did not occur on the multiple-choice section.

The ease with which one can insert humor into directions (Berk 2000, 2002, 2003) compared to content-relevant or irrelevant humor in the items should facilitate further research on this variable. Given the unexpected anxiety levels prior to the three tests, the directions hypothesis should also be tested on highly anxious examinees to give its potential moderating effect a fair test.

4.4. *Anxiety and test performance*

Another significant finding was the corroboration of the negative correlation between test anxiety and test performance with previous test anxiety research (Cassady and Johnson 2002) and statistics test anxiety research, (specifically Hong and Karstensson 2002). Higher levels of cognitive test anxiety have been consistently associated with performance decrements on course tests. Anxiety accounts for approximately 7–8% of the variance in performance. In the studies to date, the findings were based on *high cognitive* anxiety. In the current investigation, *cognitive* (PSYCH) AND *emotional* (PHYS) pretest anxiety each correlated $-.46$ with test performance on the correlation and regression multiple-choice items. That is a whopping 21% of the variance. How did that happen?

Let's examine the anxiety score distributions further. For the pre-PHYS measure, 57% of the students had 0 symptoms, 24.5% had 1, and the remaining 18.5% had 2–6 symptoms. Anxiety on the pre-PSYCH was even lower with 77.6% indicating 0 symptoms, 11.2% had 1, and the rest ranged from 2–7 symptoms. These positively skewed distributions for both subscales (PHYS $= 2.21$, PSYCH $= 3.29$) spuriously inflated their variances. Couple these distributions with the high-test performance negatively skewed distributions and the result is inflated correlation coefficients. Just how much is difficult to estimate.

4.5. *Design contributions*

Beyond the quantitative results of this study, there are elements of the design that are new and have not been addressed in previous research on

Table 5. *Contributions of this study compared to previous research on test anxiety and humor in testing*

Element	Previous research	This study
1. *Measurement of test anxiety*	State test anxiety scale: worry and emotional	State situation-specific test anxiety symptom scale: PHYS and PSYCH
a. No. of items	18–40 items	20 per subscale
b. Admin. time	5–20 min.	1–2 min.
c. Scale admin.	Days before test	Immediately before and after test
d. Test admin. format	Closed-book	Open-everything
2. *Humor in testing*		
a. Test item format	Multiple-choice	Multiple-choice, matching, and constructed-response
b. Type of course	Undergrad. psychology	Grad. statistics
c. Humorous directions	None	Humorous directions
d. Humor technique in items	Content-irrelevant	Content-relevant

test anxiety and humor in testing. A comparison of those elements is shown in Table 5. The contributions of this study focus on the characteristics of the STAS and the humor technique for three test item formats. These should provide direction for future research in these domains.

5. Conclusions

This study employed one of the most rigorous experimental designs and statistics with 80% power to detect significant between-sample differences that a researcher could use. Yet real-world teaching and testing conditions coupled with extraordinarily bright students sabotaged the results and the ability to accurately answer the stated hypotheses. The actual biostatistics course experience produced a virtual ''floor effect'' for test anxiety levels and ''ceiling effect'' for test performance. Pretest anxiety on the PHYS and PSYCH subscales was so low right before each test that it had nowhere to plummet by the end of the test. Conversely, students were so thoroughly prepared for the test that their outstanding performance had minuscule room for improvement. Although these outcomes were beneficial to the students and my course evaluations, the humorous treatment race didn't have a fair chance of winning.

Despite these limitations, seven major conclusions can be drawn from the results:

1. Humor in test directions can significantly increase test performance, particularly on constructed-response problem-solving items.
2. Cognitive and emotional situation-specific test anxiety is negatively associated with multiple-choice test performance, explaining up to 21% of the variance.
3. Cognitive and emotional dimensions of test anxiety can be measured immediately before and after test performance under real testing conditions.
4. Content-relevant humor can be integrated into several item formats (although its effectiveness was not directly evaluated by the students in this study).
5. Randomized design studies of humor in course tests can be conducted in college courses other than undergraduate psychology.
6. Well-planned teaching strategies may be more effective in reducing test anxiety (and improving test performance) than humor in the tests themselves.
7. Humor in course tests may be worthy of consideration because it poses "no harm" to performance and previous self-report studies found students prefer it.

There were so many new elements introduced in this research on test anxiety and humor in testing (see Table 5) that future research should seek to corroborate and extend this work. In fact, there are so few well-designed studies of the effects of humor on anxiety and test performance that researchers could view the preceding list of conclusions as a springboard to construct their own investigations. Furthermore, subsequent investigations should request the participants to evaluate the humor in the test to gain insight into what types of humor are funny, appropriate, and helpful in reducing anxiety and improving performance. A substantial amount of evidence is urgently needed before educators can infer that the outcomes of humor in their tests are evidence based.

The Johns Hopkins University

Note

Correspondence address: rberk@son.jhmi.edu

* The data analysis for this research was supported by funding from the Johns Hopkins University Center for Nursing Research. The interpretation of the results and the

conclusions do not necessarily reflect the views of the Center, the University, or anyone else on earth. The content should be blamed solely on the authors.

The authors gratefully acknowledge Maryke van Leeuwen and Joanna Katz for their assistance in instrument administration, scoring, and analysis. The authors also appreciate the extremely helpful comments by Jerrell Cassady, Bob McMorris, Tim Bothell, and several anonymous reviewers on a previous manuscript version of this article.

References

Abel, Millicent
2002 Humor, stress, and coping strategies. *Humor: International Journal of Humor Research* 15, 365–381.

Bar-Tal, Yoram, Alona Raviv, and Ada Spitzer
1999 The need and ability to achieve cognitive structuring: Individual differences that moderate the effect of stress on information processing. *Journal of Personality and Social Psychology* 77, 33–51.

Benjamin, Michael, Wilbert J. McKeachie, Y. Lin, and D. P. Hollinger
1981 Test anxiety: Deficits in information processing. *Journal of Educational Psychology* 73, 816–824.

Bennett, David J., and Geoffrey F. W. Turner
2001 The use of humorous alternatives in multiple-choice tests: No pain, no gain? Paper presented at the annual meeting of the International Society of Humor Studies, College Park, MD.

Benson, Jeri, and Elizabeth E. Tippets
1990 *Confirmatory* factor analysis of the test anxiety inventory. In Spielberger, Charles D., and Rogelio Dias-Guererro (eds.), *Cross-cultural Anxiety*, vol. 4. New York: Hemisphere/Taylor-Francis, 149–156.

Benson, Jeri, Marc Moulin-Julian, Christine Schwarzer, Bettine Seipp, and Nabil El-Zahhar
1992 Cross-validation of a revised test anxiety scale using multi-national samples. In Hagtvet, Knut A., and T. Backer Johnson (eds.), *Advances in Test Anxiety Research*. Lisse: Swets and Zeitlinger, 62–83.

Berk, Ronald A.
2000 Does humor in course tests reduce anxiety and improve performance? *College Teaching* 48, 151–158.
2001 The active ingredients in humor: Psychophysiological benefits/risks for older adults. *Educational Gerontology: An International Journal* 27, 323–339.
2002 *Humor as an Instructional Defibrillator: Evidence-based Techniques in Teaching and Assessment*. Sterling, VA: Stylus Publishing.
2003 *Professors Are from Mars, Students Are from Snickers: How to Write and Deliver Humor in the Classroom and in Professional Presentations*. Sterling, VA: Stylus Publishing.
2004a Coping with the daily stressors of an academic career: Try Mirthium.® *Academic Physician and Scientist* July/August, 2–3.
2004b Research critiques incite words of mass destruction. *The Humor Connection* Winter, 8–11.

(AutoPDF V7 8/8/06 08:22) WDG (148×225mm) TimesM J-1582 Humor, 19:4 PMU: C(C) 31/07/2006 pp. 425–450 1582_19-4_03 (p. 447)

Berk, Ronald A., and Joy P. Nanda
1998 Effects of jocular instructional methods on attitudes, anxiety, and achievement in statistics courses. *Humor: International Journal of Humor Research* 11, 383–409.

Cann, Arnie, Kitty Holt, and Lawence G. Calhoun
1999 The roles of humor and sense of humor in response to stressors. *Humor: International Journal of Humor Research* 12, 177–193.

Cassady, Jerrell C.
in press a The impact of cognitive test anxiety on test comprehension and recall in the absence of salient evaluative pressure. *Applied Cognitive Psychology*.
in press b Cognitive test anxiety across the learning-testing cycle. *Learning and Instruction*.

Cassady, Jerrell C., and R. E. Johnson
2002 Cognitive test anxiety and academic performance. *Contemporary Educational Psychology* 27, 270–295.

Covington, Martin V., and C. L. Omelich
1987 "I knew it cold before the exam": A test of the anxiety-blockage hypothesis. *Journal of Educational Psychology* 79, 393–400.

Deffenbacher, Jerry
1980 Worry and emotionality in test anxiety. In Sarason, Irwin G. (ed.), *Test Anxiety: Theory, Research, and Applications*. Mahwah, NJ: Lawrence Erlbaum, 111–124.

Dixon, Norman F.
1980 Humor: A cognitive alternative to stress? In Sarason, Irwin G., and Charles D. Spielberger (eds.), *Stress and Anxiety*, vol. 7. Washington, DC: Hemisphere, 281–289.

Everson, H. T., R. E. Millsap, and C. M. Rodriguez
1991 Isolating gender differences in test anxiety: A confirmatory factor analysis of the test anxiety inventory. *Educational and Psychological Measurement* 51, 243–251.

Goldstein, Jeffrey H., J. Harmon, Paul E. McGhee, and R. Karasik
1975 Test of an information processing model of humor: Physiological response changes during problem and riddle solving. *Journal of General Psychology* 92, 59–68.

Hedl, J. J., J. L. Hedl, and D. B. Weaver
1981 The effects of humor on anxiety and performance. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, CA.

Hembree, R.
1988 Correlates, causes, and treatment of test anxiety. *Review of Educational Research* 48, 47–77.

Hong, Eunsook
2001 Test anxiety, perceived test difficulty, and test performance: Temporal patterns of their effects. *Learning and Individual Differences* 11, 431–447.

Hong, Eunsook, and Lewis Karstensson
2002 Antecedents of state test anxiety. *Contemporary Educational Psychology* 27, 348–367.

Johnson, A. Michael
1990 A study of humor and the right hemisphere. *Perceptual and Motor Skills* 70, 995–1002.

Kuhlman, Thomas T.
1984 *Humor and Psychotherapy*. Homewood, IL: Dorsey Press.
Kuiper, Nicholas A., Melissa Grimshaw, Catherine Leite, and Gillian Kirsh
2004 Humor is not always the best medicine: Specific components of sense of humor and psychological well-being. *Humor: International Journal of Humor Research* 17, 135–168.
Kuiper, Nicholas A., and Rod A. Martin
1993 Humor and self-concept. *Humor: International Journal of Humor Research* 6, 251–270.
1998 Is sense of humor a positive personality characteristic? In Ruch, Willibald (ed.), *The Sense of Humor: Explorations of a Personality Characteristic*. Berlin/New York: Mouton de Gruyter, 159–178.
Lefcourt, Herbert M.
2001 The humor solution. In Snyder, C. R. (ed.), *Coping and Copers: Adaptive Processes and People*. New York: Oxford University Press, 68–72.
Liebert, R. M., and L. W. Morris
1967 Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports* 20, 975–978.
May, Rollo
1953 *Man's Search for Himself*. New York: Norton.
McMorris, Robert F., Roger A. Boothroyd, and Debra J. Pietrangelo
1997 Humor in educational testing: A review and discussion. *Applied Measurement in Education* 10, 269–297.
Morris, L. W., M. A. Davis, and C. H. Hutching
1981 Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology* 73, 541–555.
Mueller, J. H.
1980 Test anxiety and the encoding and retrieval of information. In Sarason, Irwin G. (ed.), *Test Anxiety: Theory, Research, and Appliction*, Hillsdale, NJ: Erlbaum, 68–86.
Musch, Jochen, and Arnat Broder
1999 Test anxiety versus academic skills: A comparison of two alternative models for predicting performance in a statistics exam. *British Journal of Educational Psychology* 69, 105–116.
Naveh-Benjamin, M.
1991 A comparison of training programs intended for different types of test-anxious students: Further support for an information-processing mode. *Journal of Educational Psychology* 83, 134–139.
Nezlek, John, and Peter Derks
2001 Use of humor as a coping mechanism, psychological adjustment, and social interaction. *Humor: International Journal of Humor Research* 14, 395–413.
O'Connell, Walter E.
1976 Freudian humor. The Eupsychia of everyday life. In Foot, Hugh C., and Anthony J. Chapman (eds.), *Humor and Laughter: Theory, Research, and Application*. London: Wiley, 313–330.
Onwuegbuzie, Anthony J., and Christine E. Daley
1996 The relative contributions of examination-taking coping strategies and study coping strategies to test anxiety: A concurrent analysis. *Cognitive Therapy and Research* 20, 287–303.

Perlini, Arthur H., Roger G. Nenonen, and D. L. Lind
1999 Effects of humor on test anxiety and performance. *Psychological Reports* 84, 1203–1213.

Sarason, Irwin G.
1980 *Test Anxiety: Theory, Research, and Applications*. Mahwah, NJ: Lawrence Erlbaum.
1984 Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology* 46, 929–938.
1986 Test anxiety, worry, and cognitive interference. In Schwarzer, R. (ed.), *Self-related Cognitions in Anxiety and Motivation*. Mahwah, NJ: Lawrence Erlbaum, 19–34.

Schutz, Paul A., and Heather A. Davis
2000 Emotions and self-regulation during test taking. *Educational Psychologist* 35, 243–256.

Schwarzer, Ralf
1984 Worry and emotionality as separate components in test anxiety. *International Review of Applied Psychology* 33, 205–220.

Schwarzer, Ralf, and M. Jerusalem
1992 Advances in anxiety theory: A cognitive process approach. In Hagtvet, K. A., and Thomas B. Johnsen (eds.), *Advances in Test Anxiety Research*. Lisse: Swets and Zeitlinger, 2–31.

Smith, Ronald E., James C. Ascough, Ronald F. Ettinger, and Don S. Nelson
1971 Humor, anxiety, and task performance. *Journal of Personality and Social Psychology* 19, 243–246.

Spielberger, Charles D.
1980 *Test Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Suls, Jerry M.
1972 A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. In Goldstein, Jeffrey H., and Paul E. McGhee, (eds.), *The Psychology of Humor: Theoretical Perspectives and Empirical Issues*. New York: Academic Press, 81–100.
1983 Cognitive process in humor appreciation. In McGhee, Paul E., and Jeffrey H. Goldstein (eds.), *Handbook of Humor Research*, vol. 1: *Basic Issues*. New York: Springer, 39–58.

Yovetich, N. A., J. A. Dale, and M. A. Hudak
1990 Benefits of humor in reduction of threat-induced anxiety. *Psychological Reports* 66, 51–58.

Zeidner, Moshe M.
1998 *Test Anxiety: The State of the Art*. New York: Plenum Press.

Zeidner, Moshe M., and B. Nevo
1992 Test anxiety in examinees in a college admission testing situation: Incidence, dimensionality, and cognitive correlates. In Hagtvet, K. A., and Thomas B. Johnsen (eds.), *Advances in Test Anxiety Research*. Lisse: Swets and Zeitlinger, 288–303.