Brown, Alan S.; Itzig, Jerry N.
The Interaction of Humor and Anxiety in Academic Test Situations.
[76] 27p.  HIP-$0.03 HC-$2.06 Plus Postage.
Academic Achievement; *Anxiety; *Arousal Patterns; Higher Education; *Humor; Multiple Choice Tests; *Performance Factors; Questionnaires; Research Problems; Response Style (Tests); Self Evaluation; Test Construction; Testing; *Testing Problems; *Test Items; Undergraduate Students
Self Evaluation Questionnaire; State Trait Anxiety Inventory; *Test Anxiety

ABSTRACT

The effects of humorous test questions on test performance of high and low-anxious college students was investigated. It was hypothesized that humor should reduce the anxiety level of high-anxious subjects, and thus improve their performance, while having little effect on low-anxious subjects. Students were assigned to a low or high-anxious group depending upon their responses to the Self-Evaluation Questionnaire part of the State-Trait Anxiety Inventory. Humor was introduced to the test by dropping one of the three incorrect alternatives in the multiple choice section and substituting an absurd alternative. Four experimental groups were formed by assigning low and high-anxiety groups to the humorous or nonhumorous version of the second test. High-anxious subjects did relatively better on the nonhumorous test, while low-anxious subjects scored relatively better on the humorous test. This supported the conception of an inverted U-shaped function relating arousal to performance: increased arousal induced the humorous questions facilitated low-anxious while hindering high-anxious subjects. Humorous questions should be avoided because of the risk of differentially biasing the testing situation against students who have a higher anxiety or arousal level. Two methodological problems present in previous research--initial equivalence of high and low anxious subjects on the performance dimension, and reactivity and performance-produced anxiety bias--were eliminated in the present design. (Author/CP)
The Interaction of Humor and Anxiety in Academic Test Situations

Alan S. Brown and Jerry M. Itzig

Southern Methodist University

Running head: Humor, Anxiety, and Test Performance
Abstract

The effects of humorous test questions on examination performance of high- and low-anxious subjects was investigated. It was hypothesized that humor should reduce the anxiety level of high-anxious subjects, resulting in improved performance, while having little effect on low-anxious subjects. However, the data contradicted this postulation. Comparing humorous and nonhumorous exams, high-anxious subjects performed relatively better on the nonhumorous one, while low-anxious subjects scored relatively better on the humorous one. This supported the conception of an inverted U-shaped function relating arousal to performance: increased arousal induced by the humorous questions facilitated low-anxious while hindering high-anxious subjects. Two methodological problems present in previous research were discussed, and eliminated in the present design.
The Interaction of Humor and Anxiety in Academic Test Situations

The influence of level of anxiety on test performance has been a topic of considerable interest in psychological research, from both theoretical and practical perspectives. In the theoretical sense, the anxiety-performance relationship has been used to examine internal versus external attentional mechanisms (e.g., Liebert & Morris, 1967; Wine, 1971), dominant (correct) versus nondominant (incorrect) response tendencies (e.g., Montague, 1953), and various levels of activation (e.g., Munz, Costello, & Korabik, 1975).

From a practical vantage point, the concern over the anxiety-performance relationship is obvious. Educators have long been aware of the possible detrimental effects of high anxiety levels on test performance, and, therefore, the need to eliminate its possible confounding influence. A number of different efforts have been directed at reducing the anxiety level of high anxious students, involving both direct and indirect procedures.

The direct manipulations usually involve some type of counseling, such as systematic desensitization (cf. Wine, 1971), cognitive modification (Meichenbaum, 1972), or covert reinforcement (Ouidry & Randolph, 1974). On the other hand, the indirect manipulations focus on changing the test stimulus environment as a means to reduce the level of anxiety. Within this framework, several techniques seem to have some success at reducing test anxiety (as inferred from higher test scores) for high-anxious students. These can be classed as either (a) pertaining to the general examination procedure, such as increased frequency of examinations (Dustin, 1971) and
opportunity to comment on the questions (Paul & Eriksen, 1964; Smith & Rockett, 1958), or (b) specific item manipulations, such as spreading the difficult items throughout the examination sequence (Munz & Smouse, 1968) or using humorous test items (Smith, Ascough, Ettinger, & Nelson, 1971; Terry & Woods, 1975).

The primary purpose of the present study was to attempt to reduce anxiety (and therefore improve performance) of high anxious subjects through use of humorous test questions. However, there are two important methodological problems with research on the anxiety-performance relationship, which will be commented on below. The secondary purpose of the present investigation was to present an improved procedure which will eliminate these problems and allow for a more unambiguous interpretation of the empirical outcome.

The first problem concerns the initial equivalence of the high- and low-anxious subjects on the performance, or dependent variable, dimension. As Campbell and Stanley (1966) have pointed out, it is essential to determine the equivalence of the groups at the outset of the study in order to ensure the internal validity of the results. Previous investigations in this area have consistently failed to check on the equivalence of the test-taking ability of subjects differing in anxiety level prior to the instigation of the manipulation. Although some studies have used a repeated measures design (French, 1962; Paul & Eriksen, 1964), they have not used a preliminary test as a basis for equating the groups.

The other problem relates to the procedures employed in determining the subjects' anxiety level. Typically, anxiety has been differentiated
into two different varieties: trait and state. Trait anxiety refers to an individual's general base level of anxiety which he/she typically experiences while state anxiety concerns the level of anxiety experienced in a given situation. The particular type of state anxiety of interest in the research on the anxiety-performance relationship is test anxiety. As Sarason (1960) has pointed out,

...the majority of studies relating measures of general anxiety to measures of intellectual performance have yielded nonsignificant correlations....However, studies which have related test anxiety ...to measures of intellectual performance have shown consistent negative correlations. (pp. 407-408)

Ideally, some measure of test anxiety would be more appropriate than a measure of general anxiety for studies probing the effects of a manipulation on the anxiety-test performance relationship. However, the use of a test-anxiety scale is methodologically unsuited to such studies, for the following reasons.

Reactivity bias. If the test-anxiety questionnaire is presented prior to the actual test (whether minutes, days, or weeks), there always exists the possibility that this pretest in some way sensitizes and/or alters the normal behavior of subjects on the subsequent test (see Campbell & Stanley, 1966). For instance, the fact that a student has stated earlier that he/she is high test anxious may in some way interfere with later performance, or the opposite (low anxious rating may be facilitative). It is easily conceivable that some sort of self-fulfilling expectation may mediate this effect.
Performance-produced anxiety bias. On the other hand, if the questionnaire is given after the test, it is quite conceivable that a subject may not accurately judge his/her level of anxiety during the test, but rather do one of the following:

1. Judge the current (post-test) level of anxiety and attribute it as having existed during the exam. This level of anxiety could easily be generated by the level of performance exhibited on the test. That is, successful performance would end up producing an eventual low level of post-test anxiety, while unsuccessful performance would generate a higher level of post-test anxiety.

2. Related to this first point, they may infer their level of test anxiety from their perceived performance. Specifically, they may intuitively assume that because they did poorly, they were anxious or because they did well, they were relatively at ease (disregarding, of course, their present post-test level of anxiety).

Both of these biases (reactivity and performance-produced anxiety) are technically impossible to eliminate whenever a test-anxiety questionnaire is used. Although there is no way to directly confirm the existence of these biases, indirect support is derived from the consistent negative relationship between measures of performance and test (or state) anxiety (Alpert & Haber, 1960; Cowen, 1957; Mandler & Cowen, 1958; Osterhouse, 1975; Paul & Eriksen, 1964; Sarason, 1957, 1959; Sarason & Mandler, 1952; Smith, Ascough, Ettinger, & Nelson, 1971; Snyder & Katahn, 1970; Spielberger, 1966). Whether the questionnaire is given prior to, or after, the critical performance measure, a negative relationship would be expected between anxiety
Because of these problems, a general measure of anxiety was used in the present study to differentiate between high and low anxious subjects. The use of a general scale seems to be justified because (a) previous studies show that there is a strong positive relationship between trait and test (or state) anxiety (O'Neil, Hansen, & Spielberger, 1969; O'Neil, Spielberger, & Hansen, 1969; Spielberger, Gorsuch, & Lushene, 1970; Wittmaier, 1974), (b) the test does index a general level of anxiety which persists across a variety of situations, including testing, and (c) the aforementioned biases are avoided because there is no ostensible relationship between the general anxiety scale and the testing situation.

To summarize, the present investigation involved an examination of the effects of test question manipulation (humorous versus nonhumorous) on examination performance of high- and low-anxious students, with anxiety level determined by the Self-Evaluation Questionnaire (Spielberger, Gorsuch, & Lushene, 1970) and performance levels among the four anxiety-text type groups equated on the basis of the first course examination. It was hypothesized that the humorous questions should have a facilitating effect on the high-anxious students through the reduction of examination anxiety, while having no effect on low-anxious students.

**Method**

**Subjects.** Sixty-three undergraduate psychology majors enrolled in an upper-level course (Sensation and Perception) at Southern Methodist University participated in the study.

**Group formation.** During the second week of classes, all 63 students
were given the Self-Evaluation Questionnaire, in class, in order to determine their level of trait anxiety, or A-trait (see Spielberger, 1966). The scores ranged from 20 to 70 out of a possible range of 20 to 80. The distribution was divided at the median, and the 32 subjects scoring in the low range (20 to 47) were designated as low anxious, while the 30 subjects in the high range (48 to 70) were called high anxious.

Four weeks into the semester, the first examination was given to the 63 students. Immediately after the exam, the short form of the Self-Evaluation Questionnaire (STAI Form 1) was administered, as an incidental check on the test state anxiety level. After eliminating those students who were either at a ceiling level of 98 to 100 percent (two high and two low anxious subjects) or at a basement level of 61 percent or less (two low anxious subjects), the subjects were randomly assigned to the humorous or nonhumorous condition for the second examination. This resulted in 14 subjects in each of the four experimental groups, formed by the crossing of anxiety level (high and low) and test type (humorous and nonhumorous).

Performance on the first examination was comparable across all four groups, with means of 84.54 percent for high-anxious—humorous, 83.93 percent for high-anxious—nonhumorous, 83.86 percent for low-anxious—humorous, and 82.93 percent for low-anxious—nonhumorous. The difference among these means was not statistically significant, $F(3,52)=0.04$, $p>.05$, suggesting that the groups were equivalent in terms of their initial level of test performance.

Test item construction. The second examination consisted of 25 multiple-choice questions followed by 13 short-answer questions (each section accounted for half the point total). The humor was introduced into the test in the
The interaction
8
following manner: for eight of the multiple-choice items, one of the three incorrect alternatives was randomly dropped and a humorous alternative substituted in its place (i.e., "The gelatin-like covering called the cupula fits over the: (a) crista, (b) macula, (c) sacculus, (d) pineapple chunks and Bing cherries").

It was decided to put the humor into the alternatives, rather than the stem, of the questions because (a) it would be more difficult to maintain equal difficulty of the items between the humorous and nonhumorous forms of the test if the stem was altered, and (b) the subjects could conceivably have a more difficult time separating out the humor from the essential content, in order to arrive at an answer.

**Test procedure.** The critical manipulation occurred at the second examination, given four weeks after the first. At this time, the subjects were given either the humorous or nonhumorous form of the test (as randomly determined earlier). After completion of the examination, they were asked (but not required) to fill out the Self-Evaluation Questionnaire, as after the first examination. In addition, those subjects who were given the humorous form of the test were asked to fill out a short questionnaire on the effects of the humorous questions.

**Results**

On the second examination, one subject (in the high-anxious—nonhumorous group) dropped 31 percentage points (89 to 58). It was decided to drop this subject because of the strong suspicion that he was not performing within the normal range on either or both occasions (the next largest change from the first to second examination was only 13 percent). In order to
equalize the size of the four groups, one subject was randomly dropped from each of the other three, resulting in a total of 13 subjects per group.

The average performance for each of the four groups on both examinations is presented in Figure 1. This information reveals several important points.

First, dropping a subject from each of the groups did not alter their equivalence on the first examination. A post-hoc comparison of the means (Duncan's Multiple Range Test) revealed no significant differences ($p > .05$) among them. Second, and more importantly, performance of subjects in the high-anxious—humorous condition declined while those in the high-anxious—nonhumorous group improved. This is contrary to the expectation proposed earlier that humor would improve performance for high-anxious subjects. Furthermore, low-anxious subjects also diverged on the second test, but in the opposite manner: those getting the humorous questions improved, while those getting the nonhumorous questions declined. This is also contrary to the expectation of no differential effects of humor with low-anxious subjects.

An analysis of variance confirmed the reliability of this finding, in that the interaction of anxiety level by test type by test was statistically significant, $F(1,48)=8.68$, $p < .01$. The only other statistically significant effect was due to test, $F(1,48)=6.52$, $p < .05$, with performance level increasing from the first to the second examination. All other main effects and interactions were not significant ($p > .05$).

The state anxiety levels of the four groups at the two tests are pre-
The Interaction

sentenced in Table 1. Because the questionnaire was not compulsory, there

were 12 of the 52 subjects who did not take it at one, or both, of the
examinations (these differences in N are noted in the table). Due to
this differential subject attrition, coupled with the afore-mentioned
interpretive problem when a test-anxiety questionnaire is given after a
test, the results should be cautiously interpreted. It appears that the
level of anxiety declined in all four groups. However, the drop was sta-
tistically significant for the low-anxious--nonhumorous group, only.

Comparisons were also made between state-anxiety levels and per-
formance. A Pearson product-moment correlation was computed between the
self-rated test anxiety and test score, separately for both the first
(r=-.24) and second (r=-.42) mid-term examinations. This correlation was
significant (p<.05) for the second, but not the first, exam. This finding
is congruent with a large number of previous studies noting the negative
relationship between test performance and anxiety.

Finally, the questionnaires concerning the effects of the humorous
questions on the subjects were examined. As shown in Table 2, the eval-

uations of the effects are very similar for both the high and low anxious
subjects (NOTE: because this questionnaire was not compulsory, two high
and two low anxious subjects did not fill one out). Besides the equiva-
}

lent overall subjective effects of the humorous questions on the high and
The Interaction

low anxious subjects, this information also suggests that the humorous questions (a) did tend to alleviate some of the anxiety, (b) improved the estimated level of performance, and (c) are desired on examinations.

Discussion

The most striking outcome of the present investigation was that not only did the humorous questions not help the high-anxious subjects, it even seemed to hinder performance relative to the high-anxious—nonhumor group. This, of course, is completely contradictory to the hypothesis that the humorous questions should improve test performance through a reduction of test anxiety. This result is doubly puzzling because it is counterintuitive, as well as being counterhypothetical.

An alternative possibility is that the humorous questions actually increased, rather than decreased, the anxiety level of the high-anxious subjects (inferring backwards from performance to internal state). There are, however, two lines of evidence which argue against this conjecture: (a) low-anxious subjects performed better when given the humorous test (compared to the nonhumorous test) and (b) the subjects consistently reported that the humorous questions helped to relieve their anxiety levels (subjectively assessed).

Another possible explanation is that arousal level, rather than anxiety level, may be the critical variable influencing performance in this situation. As Uehling (1972) has pointed out, it is usually assumed that there is some optimal arousal level for performance of a given task (which varies systematically as a function of task complexity, instructions, etc.). One theoretical position on the relationship between these two variables is that
"performance is related to arousal by an inverted U-shaped function, i.e., maintenance of arousal at an intermediate level will be optimal for performance..." (Uehling, 1972, p. 264) This position is supported by a number of researchers (see Duffy, 1957, and Malmo, 1957).

When applied to the present results, it is possible that high- and low-anxious subjects are distributed at different locations on an inverted U-shaped function relating arousal to test performance. Specifically, the high-anxious subjects are generally more highly aroused than the low-anxious subjects, placing them higher than (or to the right of) the low-anxious subjects on the hypothetical continuum. If the additional assumption is made that humorous questions increase the arousal (and not anxiety) level for both low- and high-anxious subjects, this would push both types of subjects to the right along the arousal continuum. The result would be an improvement in performance for low-anxious subjects (because they are at a point before, or preceding, the apex of the function) and a decrement in performance for high-anxious subjects (because they are at, or beyond, the apex of the inverted U-shaped function, already).

This interpretation fits conveniently with the obtained data: high-anxious subjects scored better when given the nonhumorous test (relative to the humorous form), while low-anxious performed higher on the humorous test (relative to the nonhumorous one). Additional support for this interpretive position can be derived from several studies. At the general level, Frankel (1969) discovered that there exists an arousal level which is optimal for emitting dominant responses, but that increases beyond this level will tend to increase the probability of nondominant responses being given, rather
than the dominant ones. This can be related to the present investigation by assuming that dominant responses are the correct ones while nondominant responses are the incorrect ones.

At a more specific level, Munz, Costello, and Korabik (1975) attempted to verify the notion that the inverted U-shaped function directly mediated the relationship between anxiety and performance in a test situation. They found that subjects with a high (debilitators) or low (non affecteds) level of arousal (or activation level) performed significantly poorer on a classroom examination than those possessing an intermediate level of arousal (facilitators). Furthermore, both Paul and Eriksen (1964) and Terry and Woods (1975) suggested that this inverted U-shaped conceptualization provided a good interpretive framework for their results, where attempts were made to directly manipulate anxiety levels in the test situation. This held true for both adults (Paul & Eriksen) and children (Terry & Woods).

Before finishing a discussion of this result, it should be contrasted to Smith et al. (1971), because they also used test question humor in an attempt to manipulate anxiety level in a regular examination situation. Their outcome appears to contradict the present findings, in that humorous items facilitated performance for high-anxious subjects while having no effect on low-anxious subjects. However, it is difficult to directly compare the two studies because their study suffers from the two methodological flaws which the present study attempted to correct (i.e., determination of anxiety through a test-anxiety questionnaire and no equating of base level performance for the various groups). It is interesting to note, however, that their moderate-anxiety subjects exhibited a substantial (but not sig-
significant) decrement in performance when given the humorous test (relative to the nonhumorous form). This forced them to concede that "...the manner in which humor affects performance may be a complex function of the level of anxiety being experienced." (p. 245) They also include an additional observation, which is even more pertinent to the present interpretive position: "several of the subjects in the moderate-anxiety—humorous-test condition later remarked that while the humorous items served to 'loosen them up' somewhat, they also had a somewhat distracting effect on them." (p. 245) Similar to this, the high-anxious subjects in the present study may have had their subjective level of tension reduced (as suggested by the questionnaire) while, at the same time, experiencing an increase in their arousal level (which they were not directly aware of, but which nonetheless created a distinct hindrance to performance).

Several other aspects of the results deserve comment. To begin with, both the lack of a relationship between general anxiety level and test performance (no significant difference among groups on the first examination) and presence of a negative relationship between test-state anxiety and test performance are congruent with previous studies, as discussed earlier. Further, there was a positive (but nonsignificant) correlation between trait and test-state anxiety levels for both the first (r=.27) and second (r=.23) examinations. This is also consistent with previous research, and gives some additional post hoc justification for using the general anxiety measure as a valid index of test anxiety.

Conclusion

The present investigation was an attempt to approach the investigation
of the anxiety-performance relationship in a practical setting with the use of improved methodological techniques. In terms of the improved methodology, it was suggested that there is no way to obtain an unbiased indication of test anxiety directly through a test-anxiety questionnaire, and that a general anxiety measure is preferable. In addition, some sort of performance premeasure is essential for interpretive clarity. The outcome of the study suggested that an attempt to reduce the anxiety level of high-anxious subjects through the use of humorous test items was not successful. In fact, it appears that the humorous questions increased the arousal level of both the high- and low-anxious subjects, a shift which was beneficial for the low-anxious but detrimental for the high-anxious subjects.

The practical implications of this finding are clear. It is almost a universal assumption that humor in examinations will in some way aid the students taking the test. However, the present study indicated that this may not be the case, especially when considering high-test-anxious students (see Levine and Abelson, 1959, for more discussion of this point). Therefore, it is suggested that humorous test questions should be avoided because of the risk of differentially biasing the testing situation against students who have a higher anxiety, or arousal, level.
References


Duffy, E. The psychological significance of the concept of "arousal" or activation. Psychological Review, 1957, 64, 265-275.


Footnotes

The authors wish to express their appreciation to Dr. William H. Tedford for permission to use his class for the study, as well as for construction of examination questions.

1Thanks are due to Dr. James Papy for his permission to use the retrospective version of the Self-Evaluation Questionnaire.
Table 1
Mean State Anxiety Levels at Both Examinations for each Experimental Group

<table>
<thead>
<tr>
<th>Group</th>
<th>Examination 1</th>
<th>Examination 2</th>
<th>t ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-anxious--humor(^a)</td>
<td>62.88</td>
<td>59.88</td>
<td>1.26</td>
</tr>
<tr>
<td>High-anxious--nonhumor(^b)</td>
<td>61.10</td>
<td>57.90</td>
<td>0.91</td>
</tr>
<tr>
<td>Low-anxious--humor(^b)</td>
<td>60.00</td>
<td>56.60</td>
<td>1.81</td>
</tr>
<tr>
<td>Low-anxious--nonhumor(^c)</td>
<td>58.92</td>
<td>54.75</td>
<td>2.55*</td>
</tr>
</tbody>
</table>

\(^a\) N=8
\(^b\) N=10
\(^c\) N=12
* \(p<.05\)
### Table 2
Response Frequencies to Questionnaire
Assessing the Effects of the Humorous Questions

<table>
<thead>
<tr>
<th>Question</th>
<th>High-anxious subjects</th>
<th>Low-anxious subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. When you encountered a humorous question, how did it make you feel:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. tense</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>b. relieved</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>c. no different</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2. Did the humor:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. facilitate</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>b. hinder</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>c. have no effect on your answer to the question it was in?</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3. Do you feel that the humorous questions:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. helped</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>b. hindered</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>c. had no effect on you while you were taking the test?</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
4. Did you find the humorous questions:
   a. easier  
   b. harder  
   c. no different  
   to understand than the nonhumorous questions?

5. Would you like to see tests contain:
   a. more  
   b. less  
   c. about the same number of humorous questions as this test?

\[ N=11 \]
Figure Captions

Figure 1. Mean percent correct on the first and second examinations for each experimental group.