A determination of the efficacy of a number of techniques for reducing the debilitating effects of test anxiety during university mid-term examinations.

Richard Kenneth Wallin
University of Massachusetts Amherst
A DETERMINATION OF THE EFFICACY OF A NUMBER OF TECHNIQUES FOR REDUCING THE DEBILITATING EFFECTS OF TEST ANXIETY DURING UNIVERSITY MID-TERM EXAMINATIONS

A Dissertation Presented

By

RICHARD KENNETH WALLIN

Submitted to the Graduate School of the University of Massachusetts in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY


Educational Psychology
A DETERMINATION OF THE EFFICACY OF A NUMBER OF TECHNIQUES FOR REDUCING THE DEBILITATING EFFECTS OF TEST ANXIETY DURING UNIVERSITY MID-TERM EXAMINATIONS.

A dissertation presented by RICHARD KENNETH WALLIN

Approved as to style and content by:

James M. Royer, Chairperson of Committee

Beth Sulzer-Azaroff, Member

J. Gregory Olley, Member

C. Kay Smith, Member

Jerome Myers, Department Head
Psychology
Three techniques for alleviating test anxiety effects were tried. These were: (a) playing soft background music during the exam, (b) allowing the students to make comments about the items on the exam on a separate sheet during the exam, and (c) having the exam proctor disclose to the students that he felt anxiety during tests and that he managed to overcome this anxiety by minding four hints, which hints were then distributed to the students. Test anxiety was measured using the Alpert-Haber scale and effects on performance were judged by examining scores on two mid-term exams in introductory psychology, each of which was composed of 50 multiple choice items. No statistically significant effects were noted for any of the techniques. It was concluded that none of the techniques were effective at reducing test anxiety effects during normal college exams although it was considered possible that the second of these techniques listed above would be effective if the comments were solicited between items on the examination sheets themselves.
Table of Contents

Introduction .............................................. 1.
Method .................................................. 46.
Method - Supplementary Experiments ............... 52.
Results .................................................. 59.
Results - Supplementary Experiments ............... 76.
Discussion .............................................. 85.
References ............................................. 98.
Appendices ............................................. 105.
<table>
<thead>
<tr>
<th>Table Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means of Experimental Variables by Group</td>
<td>61</td>
</tr>
<tr>
<td>One Way Analyses of Variance of Anxiety Measures by Groups</td>
<td>62</td>
</tr>
<tr>
<td>One Way Analyses of Variance of Performance Measures By Groups</td>
<td>63</td>
</tr>
<tr>
<td>Correlations Between Anxiety Measures and Performance Measures</td>
<td>64</td>
</tr>
<tr>
<td>Analyses of Variance: Performance Scores by Group and Anxiety Measures</td>
<td>66</td>
</tr>
<tr>
<td>Breakdown of Performance Scores by Group and by High, Medium, and Low Anxiety</td>
<td>68</td>
</tr>
<tr>
<td>Analyses of Variance: Scores on First Exam for Item Types Analysed by Group and Combined Anxiety Score</td>
<td>72</td>
</tr>
<tr>
<td>One Way Analyses of Variance of Questionnaire Data by Groups</td>
<td>74</td>
</tr>
<tr>
<td>Means of Subject Measures by Group</td>
<td>78</td>
</tr>
<tr>
<td>t-Tests Comparing Control Group Subjects versus Other Subjects on All Anxiety Measures</td>
<td>79</td>
</tr>
<tr>
<td>Correlations Between Anxiety Measures and Performance Scores</td>
<td>80</td>
</tr>
<tr>
<td>Means of Subject Scores by Group</td>
<td>82</td>
</tr>
<tr>
<td>t-Tests Comparing Control Group Subjects with All Other Subjects on All Anxiety Measures</td>
<td>83</td>
</tr>
<tr>
<td>Correlations Between Anxiety Measures and Examination Score</td>
<td>84</td>
</tr>
</tbody>
</table>
Introduction

The purpose of this study will be to explore a number of alternative methods for alleviating the effects of test anxiety in a normal college testing situation. In this introduction the causes and effects of test anxiety will be examined, the testing situation changes will be detailed and relevant research cited, and a rationale for the experimental design and measures will be presented.

Test anxiety is the term given to the specific debilitating anxiety experienced by a variety of people when they are put into a testing situation. Occasionally, such anxiety translates into severe performance problems in which students are unable to perform at even the most minimal levels. These are the people who 'freeze', 'choke', or 'fall to pieces' at exams. More often, however, test anxiety causes less dramatic decrements in performance which are, nonetheless, both statistically significant and of practical importance. The fact is, people who have test anxiety score lower on tests which are evaluative in nature, such as intelligence tests and academic achievement tests. This primary effect has been demonstrated repeatedly for students from early elementary school age (for example, Hill & Sarason, 1966; S. Sarason, Hill, & Zimbardo, 1964), right up to the college level (for example, Alpert & Haber, 1960; I Sarason, 1961). It is this debilitating effect that raises the most interest in test anxiety, and which will be the focus of this study.

Test anxiety is also related to decrements in performance on a variety of laboratory tasks. Paired associate learning (e.g., Mendelson, 1973), simple math problems presented by computer (Tobias, Hadl, & Towle, 1974),
serial list learning (I. Sarason, Pederson & Nyman, 1968), digit symbol learning (Milford, 1973; I. Sarason & Palola, 1960), and concept acquisition (Tennyson & Boutwell, 1973) are examples of such tasks that have been investigated and on which subjects have shown performance decrements correlated with their test anxiety.

Interestingly, test anxiety has also been shown to correlate with a complex of personality factors. In children:

the evidence suggests that highly test anxious children: (1) are self-disparaging; (2) are unadventurous; (3) possess more negative personality characteristics; and (4) have a strong tendency to indulge in daydreams. Classmates appear to react unfavourably against the high-anxious, while teachers, after the first few years at least, see them as possessing characteristics currently regarded in western culture as negative and unfavourable. Fathers....tend to view their offspring in the same way as the children judge themselves. (Gaudry & Spielberger, 1971, p.22)

In adults, it has been shown that high test anxiety correlates negatively with "a questionnaire designed to measure the subjects' expectations of achieving valued goals" (Strassberg, 1973) and that high test anxiety correlates negatively with optimism and positively with neuroticism (Walsh, 1968). Also, it has been shown that people with high test anxiety are more self-preoccupied and less cognizant of cues (Wine, 1971). These should be sufficient to emphasize the fact that test anxiety is a wide ranging and complex phenomenon.

While the decrements listed above are well documented, their causes are of uncertain origin, at best. There have been a number of serious attempts to analyze test anxiety in theoretical fashion and these will be detailed below. First, however, it might be profitable to examine
some of the positions that have developed over the years as theorists have attempted to describe and explain more general forms of anxiety. It will then be possible to relate the more specific comments on test anxiety to these more general theories of anxiety.

**Theories of General Anxiety**

Real interest in anxiety as a fundamental human condition was first stimulated by the writing of Freud. One of the reasons that his writing was so widely discussed was that, in his view, anxiety was hardly a unitary concept. He identified a number of different kinds of anxiety and postulated a variety of mechanisms to explain their existence.

Freud's three basic forms of anxiety are reality anxiety, moral anxiety, and neurotic anxiety. Neurotic anxiety, the kind to which he directed most of his attention, could be displayed in three forms, free-floating anxiety, phobic anxiety, and the panic or near panic state of an anxiety attack. Of these, his definition of phobic anxiety is the one which, in some cases, would seem to best describe test anxiety. However, the mechanisms postulated by Freud for explaining neurotic anxiety would seem to be inappropos when applied to test anxiety. Depending on when in his career he was writing, Freud believed either that the anxiety was produced by the repression of sexual impulses or that the anxiety was caused by the conflicts between the ego, the id, and the superego, which were brought on by the id instincts (and which then caused the repression.) In either case, the anxiety was a by-product of the id instincts toward sexual gratification. This basis would seem to be inappropriate as a link with test anxiety.

One is tempted to relate reality anxiety to test anxiety. Reality
anxiety is based upon perception of a dangerous condition in the external world. Freud related this type of anxiety to, what he calls, 'primary anxiety', anxiety which is modeled upon the process of birth, and to 'separation anxiety', anxiety which relates to any major separation of mother and child, the most traumatic of which is, of course, the trauma of birth. This type of anxiety has four constituent factors:

"(a) the flooding and overwhelming of the mental apparatus with excitation;
(b) the passivity and helplessness of the organism;
(c) the existence of separation fears that correspond to the actual physical separation of the fetus from the mother; and
(d) the autonomic quality of the organism's affective experiencing."

(Fischer, 1970, p.9). (It should be noted that later writers have pointed out that young animals who are actually experimentally separated from their mothers do not react in a way which looks consistent with the notions of anxiety expressed above. Nonetheless, the idea of 'primary anxiety', caused mainly by the trauma of birth, might still be valid as long as one deemphasizes the role of 'separation' in the origin of the 'primary anxiety'.)

According to Freud, any situation of later life that threatens to reduce the person to a state of infantile helplessness will arouse a reality anxiety signal. It will be seen later that the four factors listed above, and this sense of 'infantile helplessness', fit neatly into one or another of the theories of anxiety. However, in Freud's work, there is no explanation offered for the fact that one person will perceive a situation (such as a test) as a threat, whereas another person won't, although he does lean to some sort of unspecified environmental learning experience.

Although Freud's writing lacks a unitary concept of anxiety, it can
be said that each of his types of anxiety and each of the factors that makes them up, is fairly concretely defined. In marked contrast to this concreteness are the anxiety theories espoused by Rogers, May, and Kierkegaard. Rogers and May both agree that feelings of uncertainty and helplessness are involved in anxiety, but rather than postulating a relationship of these feelings with separation anxiety or birth trauma, they each, in their own way, postulate traumas of a more cognitive and humanistic variety. Rogers (1951) believes that anxiety is experienced when the individual perceives something that is a threat to his self-concept. May (1950) hypothesizes a source that sounds very similar to Rogers', except that May refers to a threat to the core or essence of the personality rather than to self-concept. Also, May goes farther than Rogers and states that anxiety is the response when the source of the threat is unclear, and is, in this way, following in the Freudian tradition of unconscious neurotic fears.

Kierkegaard, as an existentialist, postulates perhaps the most ephemeral cause of anxiety. Kierkegaard begins with the idea that man cannot be understood empirically, that is, he does not conceive man to be simply a passive recipient of stimuli or merely a respondent to his environment. Man and environment, in Kierkegaard's view, are co-defining, and he believes that through his choices an individual makes both himself and his world. Fischer (1970) has shown how Kierkegaard relates this view of man with anxiety. In his opinion, Kierkegaard understands anxiety to be:

an experiential state, constituted by the individual's awareness of his own possibilities, by his realization that he has no objective justification for choosing among them, and by his limited capacity to foresee all the consequences of a possible choice. To complicate
matters further, the individual may realize that if he enacts certain choices, he will be individualized as a separate entity, potentially in conflict with his fellow citizens and personally responsible for all the consequences. On these grounds anxiety inevitably occurs.... (Fischer, 1970, p.9)

Whether this mechanism can logically be applied to some people's experience in exam situations is, of course, an open question.

At the other end of the spectrum from these ephemeral conceptions of anxiety is the position adopted by the behaviourally inclined conditioning theorists. Their position is that anxiety is a form of conditioned fear, a form in which the source of the fear is vague and obscured (Dollard & Miller, 1950). In the classic "little Albert" experiment, Watson demonstrated that fears could be conditioned and generalized, and this has been adopted by most behaviourists as the mechanism by which phobic fears (such as the fear of tests) come into existence. Because they are willing to accept the fact that the source of such an anxiety is unknown, conditioning theorists do not bother to hypothesize likely causes, that is, they don't bother looking for the stimuli that correspond to the loud noise presented in the "little Albert" experiment. Still, as will be discussed later, this theory is the one on which most psychotherapeutic treatment of test anxiety is based, and the success of this treatment augers well for the validity of the theory.

Aside from the conditioning theorists, there are some behaviourists who look to work done by Pavlov to explain anxiety. In Pavlov's experiments he was able to induce experimental neuroses by confronting his animals with conditions such as: (a) intense primary stimulation, such as loud noises; (b) difficult discriminations: and, (c) conflict between excitatory and
inhibitory tendencies directed toward the same object; and others. It is possible to describe a test in such a way that it would fit into any or all of these conditions. What is difficult using only this data is to discriminate those people who will be affected by these aspects in tests from those who will not. This is a serious limitation on the use of this theory to explain test anxiety.

An interesting, though limited, view of anxiety is expressed by Mandler and Watson (1966) who postulate that interruption of behaviour will cause anxiety. Their theory is that if the interruption is unexpected the individual will become somewhat anxious and will cast about looking for an alternative path. If he is unable to find such an alternative he will become more anxious as he succumbs to a feeling of helplessness. The relationship of this theory to all those who postulate helplessness as an important ingredient in anxiety is obvious. Also, it is possible to see a relationship between their concept of interruption and the concept, to be presented below, that anxiety is in part dependent on there being no obvious method for avoiding a threatening situation. Unfortunately, there is no research which shows that interruptions will heighten test anxiety or its resultant performance decrements.

Another theorist who stresses the multidimensionality of anxiety is Carroll Izard. In his work (Izard, 1972) he tries to show that anxiety is a variable combination of two or more of what he considers to be the fundamental emotions or their components. In particular, he proposes that anxiety involves fear and two or more of the emotions of distress, shame, guilt, anger, interest, and excitement. He goes on to show that the definitions of anxiety presented by most other theorists in the field
include a number, but not all, of these emotions. It seems that his techniques allow for one to determine just what is happening emotionally to a person in a particular anxiety provoking situation, with considerable accuracy and with a great deal of specificity. However, his theories do not explain either the causes or the behavioural results of any particular anxious experience, and so have a limited ability to explain or describe the mechanisms of test anxiety.

The anxiety theorist whose work is perhaps the most compatible with a discussion of the theories of test anxiety is Seymour Epstein. In his work (Epstein, 1972) he discusses the problem of understanding anxiety from a number of different viewpoints and is quite successful at combining these into a manageable few. First, in a review of the literature, he concludes that there are three basic sources of anxiety, namely, 'primary overstimulation', 'cognitive incongruity', and 'response unavailability'. He then proceeds to show how these three basic sources fit into a model that he has developed which describes the conditions necessary for producing 'high diffuse arousal'. In a separate section, he uncovers a number of the causes of such 'high diffuse arousal'. He shows that in addition to unawareness of the source of threat, one of the causes of 'high diffuse arousal' is awareness of the source of threat but unawareness of or the impossibility of purposeful activity to avoid or reduce the anxiety causing stimulus, or the lack of an obvious avenue of flight. He also discusses the relationship of arousal to awareness, showing that at relatively low levels of stimulation an orderly expansion of awareness occurs, whereas at high levels there is a defensiveness against stimulation and this defensiveness cuts down awareness, except as it is modulated by habituation.
It will be seen that most of the facets of this theory blend well with the theories and research of test anxiety. In particular, Epstein's concept of 'high diffuse arousal' will be referred to often in the pages that follow.

**Trait versus State Anxiety**

There is one fairly separate theoretical issue which should be mentioned in conjunction with a listing of the various theories of anxiety. Some theorists feel that the reason discussions of anxiety are so confusing for the lay reader is that there are really two kinds of anxiety, namely, 'trait anxiety' and 'state anxiety'.

'State anxiety' is the term used to refer to:

> the complex emotional reactions that are evoked in individuals who interpret specific situations as personally threatening. If a person perceives a situation as threatening, irrespective of the presence of real (objective) danger, it is assumed that he will respond to it with an elevation in state anxiety, that is, he will experience an immediate increase in the intensity of an emotional state characterized by feelings of tension and apprehension, and by heightened autonomic nervous system activity. The intensity and duration of this state anxiety reaction will be determined by the amount of threat that is perceived, and by the persistence of the individual's interpretation of the situation as dangerous. (Spielburger, 1972, v.1, p. 30)

Accordingly, the test anxious person can be said to be suffering from 'state anxiety' when he is actually showing symptoms of anxiety when in a testing situation, or when one is impending.

'Trait anxiety', on the other hand, refers to:

> relatively stable individual differences in anxiety proneness, that is, to differences in the disposition to perceive a wide range of stimulus situations as dangerous or threatening, and in
the tendency to respond to such threats with state anxiety reactions. Trait anxiety may also be regarded as reflecting individual differences in the frequency and the intensity with which state anxiety (responses) have been manifested in the past, and in the probability that such states will be experienced in the future. Persons who are high in trait anxiety tend to perceive a larger number of situations as dangerous or threatening than persons who are low in trait anxiety, and to respond to threatening situations with state anxiety elevations of greater intensity.

(Spielberger, 1972, v. 1, p. 39.)

It should be noted that this definition of trait anxiety in itself contains a description of two different types of people, and this difference becomes important in a discussion of test anxiety. The first kind of person who suffers from trait anxiety is the person who "perceives a wide range of stimulus situations as dangerous or threatening." This is the person who is almost constantly anxious, as the most innocuous situation is likely to be seen as threatening. On a testing day such a person might perceive a dozen different situations as anxiety provoking, situations such as driving to school, finding the exam room, checking his watch to see if he is on time, and others, in addition to the exam itself. The second type of trait anxious person is the person who tends to "respond to threatening situations with state anxiety elevations of greater intensity." Clearly, in a testing situation, such a person will respond with an unusually high level of state anxiety. Note that, over time, both these types might have equal levels of trait anxiety, that is, the summation over time of their state anxiety levels will be equal. However, even though they have equal levels of (general) trait anxiety, they may not have equal levels of test anxiety, and it can be seen that they might respond in somewhat disimilar fashion to changes in testing environments, particularly those designed to
reduce the decrements caused by test anxiety. First, one might suppose that a person who is almost always in an anxious state will grow accustomed to it and will not suffer the decrements in performance that would be seen in the other type of trait anxious person. Also, such a person would be helped by a condition that was calming, in a general manner, such as music. On the other hand, a person whose response to a threatening stimulus is often exaggerated might be helped if the exam lacked the specific threatening stimuli to which he responds. This person would probably perform quite well at a task if it were labelled a 'game' rather than a test, or if he thought the results would not be reported.

One remaining definitional problem is that, having now distinguished between trait and state anxiety, confusion still remains when one tries to apply these concepts to test anxiety. This confusion remains because test anxiety is not one or the other of these types, but is actually both. That is, test anxiety is the predisposition carried around by the individual to react with a high degree of state anxiety in testing situations. Consequently, as will be detailed below, there are both trait and state measures of test anxiety.

Theories of Test Anxiety

Having examined some of the more prominent theories of general anxiety, we will now examine the relationship of these theories with the theories that have been developed to explain test anxiety in particular. First, however, it must be noted that in one major respect, test anxiety is defined somewhat differently that general anxiety. Most researchers either define anxiety to be occurring when the person displays certain physiological symptoms (such as palpitation, sweating, tachycardia, pallor, urinary
frequency, vertigo, headache, chest pain, anorexia, nausea, abdominal cramps, tremors, or weakness; Branch, 1965), or they will accept a person's self-report to be a valid indication of their being in an anxious state. In contrast, researchers will usually define test anxiety to have occurred when a person displays the performance decrements in a testing situation that have been outlined above. Self-reports are still considered valid, but only to the extent that they indicate a behavioural change. (Note that the occasional researcher will accept 'facilitating anxiety' as a valid form of test anxiety. This is still a behavioural change, even though it isn't a decrement.)

This difference is most evident when one examines the thrust of most theories of test anxiety. For the most part they do not attempt to find causes for test anxiety or to describe the underlying emotions involved. Instead, they focus on more detailed descriptions of what exactly the person is doing which is causing him to perform at a lower level on a task than might have been otherwise expected.

The most widely quoted attempt to describe these other behaviours is one done by Wine (1971). In her article, Wine reviewed much of the literature of test anxiety and distilled out five major facts: (a) that test anxious people are more self-preoccupied; (b) that self-focusing is activated during tests; (c) that conditions of achievement or ego-involvement are most important for arousing test anxiety; (d) that high test anxiety reduces the subject's use of cues; and, (e) that worry is a more important factor than emotionality. She then combined these facts into a theory which says that the reason highly test anxious people do poorly on tests is that they spend a large amount of their test time attending to personal,
self-preoccupational cues.

In an attempt to validate her findings, Wine attempted to train subjects to attend to their tasks rather than to themselves during testing situations. Her two major finding were that her treatment: (a) reduced the degree to which her subjects reported experiencing anxiety, and (b) increased their levels of performance on several different tasks. Unfortunately, due to certain methodological faults in her design, one must reserve until replication judgement on the efficacy of her treatment methods. Nonetheless, we may speculate on the relationship of her theory to the ones described above.

First, it is interesting to note that people who are test anxious are also self-preoccupied. This may be taken as support for the positions of May and Rogers that testing is perceived as a threat on the person's view of himself and would elicit self-focusing behaviour. Also, her finding, that training her subjects to behave appropriately in tests led to a lessening of their anxiety, is consistent with Epstein's observation that anxiety is produced by a lack of an appropriate purposeful activity to reduce the perceived threat. In this case, by providing her subjects with such an activity (even though it doesn't rationally seem to reduce the perceived threat), she effectively changed their reaction from one of anxiety to one lacking in anxiety.

In a somewhat different approach, I. Sarason (1972) attempted to show that what distinguishes the highly test anxious individual is: (a) the manner in which he attends to the events of his environment, and (b) how he interprets and utilizes the information provided by these events. He focuses his remarks on cue utilization and attentional processes.
However, when all is said, he comes to the conclusion that test anxious persons, when presented with cues that suggest that their behaviour will be evaluated, plunge inward and (a) neglect or misinterpret informational cues that may be readily available to them, or (b) experience attentional blocks. It can be readily seen that this is basically Wine's position with the addition of the attentional blocks. These attentional blocks correspond rather neatly with Epstein's comments on the fact that 'high diffuse arousal' will cause a constriction in awareness in order to cut down on overstimulation.

Mandler (1972) discusses the exact same behaviour, and attempts to show that the crucial variable involved is 'self-instruction'. That is, Mandler suggests that the reason students become self-preoccupied and do not attend to task oriented cues is because they, mistakenly, instruct themselves to do so. In other words, according to Mandler, the highly anxious student, put into a testing situation, will say to himself: "This is the way I usually behave in a testing situation, and therefore, this is the way I should behave." He then goes on to show that the empirical support used by both Wine and Sarason actually support his position.

Of course, what he is proposing is an intermediate mechanism, one which requires more complex covert cognitive behaviour that either of these two other explanations, and unfortunately, Mandler presents no data to support his supposition in particular.

There are a number of theorists who attempt to divide test anxiety into constituent factors. Sassenrath (1964) did a factor analysis of the items on S. Sarason's Test Anxiety Questionnaire (which will be discussed in detail later in this introduction). He came up with seven first order factors.
Gorsuch (1966) reanalyzed Sassenrath's data and came up with two second order factors, which he labelled 'emotionality' and 'anxious avoidance of testing'. Liebert and Morris (1967; Morris & Liebert, 1969) used these analyses to suggest that test anxiety was, in fact, composed of two factors, worry and emotionality. They defined worry to be the cognitive factors relating to test anxiety, mostly lack of confidence, and emotionality to be the autonomic factors and reactions that tend to occur under exam stress. In their first study, they tried to show that worry would vary inversely with the subjects' expectation of success, but that emotionality would be highest when the subjects were not sure how they would do, whether good or bad. In fact, worry varied as they had hoped it would, but emotionality did not vary at all between groups of various levels of expectancy of success. In their second study they tried to manipulate emotionality, as they had defined it. They divided their subjects into two groups, one of which got a test in a 'timed' condition, while the other group had unlimited time in which to take their test. As will be discussed later, timing a test generally results in an increase in test anxiety effects. Their prediction was that this would indeed happen, and that these effects would vary with emotionality. In fact, the effects that did appear varied with worry and not with emotionality. That this happened suggests that there are interesting distinctions to be drawn between separably factors, but that, as yet, describing these factors with any sort of precision is not possible.

Another group of theorists have concentrated on a different aspect of the problem of dividing test anxiety into its constituent factors (Munz, Costello & Korabik, 1975; Sweeney, Smouse, Rupiper & Munz, 1970).
Using the Alpert-Haber Achievement Anxiety Test (to be described later) they divided their subjects into four groups. On the AAT subjects can score either high or low in either 'debilitating anxiety' or 'facilitating anxiety'. Thus, the four groups that they described were: 'debilitators' (those subjects scoring high on debilitating anxiety but low on facilitating anxiety), 'facilitators' (those scoring low on debilitating anxiety but high on facilitating anxiety), 'high-affecteds' (those scoring high on both scales), and 'non-affecteds' (those scoring low on both).

As a first step, they showed that facilitators and debilitators view their own internal condition in an exam quite differently. Using the Activation-Deactivation Adjective Check List, they found that debilitators perceived themselves to be mostly in a state of 'high activation' which corresponds to a stress type reaction. On the other hand, facilitators scored themselves more often in a separate dimension labelled 'general activation' which is a more peppy or lively type of activation.

Having identified separable groups, they reasoned that they might affect performance differentially by varying activation levels in their subjects. They reasoned that if they increased activation they would hurt debilitators but help facilitators and if they decreased activation the opposite would happen. To increase activation, they arrived late for an important exam, told the students it was very hard, and were nasty and unpleasant. To decrease activation they allowed their students to make comments on the items in the exam. (This technique of allowing commenting will be described in more detail later.) As it turned out, only the debilitators were affected by the conditions, and they were only affected (to their advantage) by the low activation condition where commenting was allowed.
The relationship between activation level and either its facilitative or debilitating effect is an issue of some importance theoretically. Munz and his associates were trying to explain what is called the inverted U phenomenon (Malmø, 1957). The inverted U is the shape of the graph of the relationship postulated to explain the effects one gets when varying activation level or arousal level in test anxiety studies. According to the inverted U hypothesis subjects will perform at their highest level if they are in an intermediate level of arousal or anxiety. That is, if they experience no anxiety, or they experience very great anxiety, their performance will not be as good as if their anxiety is at a medium level. An example of this relationship can be seen in an experiment done by Sarason in 1972. In this particular experiment Sarason varied the instructions given to his subjects before they were to undertake a verbal learning task. If he gave them anxiety provoking instructions, that is, instructions in which he emphasized the evaluative nature of the task, then people with high test anxiety did much worse than people with low test anxiety. On the other hand, if he gave them very reassuring instructions in which he emphasized that he just wanted to see how people responded in the task situation, people with low test anxiety did much worse than people with high test anxiety.

It can be seen that the inverted U hypothesis is consistent with the nature of anxiety postulated by Epstein. As has been detailed above, Epstein tried to show that relatively low levels of stimulation an orderly expansion of awareness occurs while at high levels of stimulation there is defensiveness against it and this cuts down on awareness. This position is at variance with that of Spence and Spence (1966) who tried to show
that anxiety effects were consistent with the effects of a rise in drive (or arousal) level and that these effects were a function of a linear (multiplicative) relationship between habit strength and drive (that is, \( E \) (excitation, likelihood to respond) = \( H \times D \) (habit strength times drive)).

To explain the effects one normally gets in test anxiety studies, Spence and Spence postulate the existence of separate task irrelevant responses which become predominant in high anxiety test situations. Possible confirmation of their position can be taken from a further experiment done by Sarason (1972) in which he managed to invent an instruction that raised the level of correct responding of subjects with both low and high test anxiety.

While trying to demonstrate that the results shown above can be interpreted in a way which is consistent with their theory (that is, one that postulates a monotonically increasing effect of anxiety on performance), Spence and Spence also put forward the notion that it might be possible to postulate a nonmonotonic relationship between drive level and the experimental variables determining it. Of course, if you are willing to postulate such a relationship, there is no reason not to suppose that such a relationship would interact with a non-monotonic relationship between drive and performance in such a way as to explain Sarason's experimental results.

It is also possible that in his special instructions Sarason was speaking, as it were, to only one kind of person, the highly test anxious kind. That is, his instructions might only have reduced anxiety (drive, arousal) in certain kinds of people, namely, highly test anxious people. If this were true, then one could still hypothesize an inverted U relationship
between anxiety level and performance and account for all differential effects of instructions.

With this perspective it is interesting to examine the results obtained by Munz and his colleagues. If they could have shown a separable relationship between facilitation and debilitation, then they would have been in a position to discredit the simple model of the relationship between arousal and performance. However, they failed in two ways. First, as has already been mentioned, they couldn't get a lowering of the performance of their low anxious subjects. More important, they got very confusing results from those students who scored high on both facilitation and debilitation, so confusing in fact, that they don't even bother reporting their data on these people. This must indicate that there are complications involved in the relationship between arousal and anxiety and facilitative and debilitative effects, and that these complications do not fit neatly into their theoretical model.

Test Anxiety as a System

Given the large number of theories and explanations for the effects of test anxiety, one wonders whether it is likely that one of them will come to be accepted as valid, or whether it would be possible to integrate some of them into a single, coherent, all-inclusive, theory. It is the opinion of this writer that it will not be possible to construct a theory which would account for all the effects that have been demonstrated to be related to test anxiety. Rather, one must accept the fact that test taking behaviour is a complicated activity and that in order to understand it one must be willing to recognize that many different mechanisms are operating
inside the individual involved and that it is the summation of all these separate effects which causes the overall decrement which has been correlated with test anxiety. In this section I will attempt to show how a coordinated view of test anxiety can include most of the different mechanisms postulated by the theorists reviewed above and make more sense doing so than an artificial, albeit parsimonious, theory.

It is perhaps easiest to visualize such a coordinated view as a complex system at work. As with any system, there will be steady-state behaviour, there will be change over time, and there will be the occasional counter-intuitive effect.

When analyzing a system, one normally starts by identifying its end result. In this case the one central behavioural index which shows the condition of the system is the ability or tendency of the subject, at any instant, to choose the correct answer from among alternatives on a test or test-like task. In any test the final score or total performance measure will be the summation of the number of these correct responses.

It is important to note that the ability of the subject to choose the correct alternative varies from question to question. Consequently, we will examine first the mechanisms operating on the choice process at any one point in time and then will show how these mechanisms operate to change the total likelihood of a correct response from question to question.

In order to answer a question correctly (assuming for the moment that the student does in fact have that information stored) the student must do three things. He must read the question, he must search his memory for the correct answer, and then he must choose among alternatives (whether examiner constructed, or self-constructed as in an essay exam) the alternative which
most closely resembles his perception of the correct answer. Each of these behaviours is differentially affected, in turn, by a number of internal conditions. Two of these internal conditions are the general arousal or drive level, and the inappropriate focusing behaviour level. We will discuss here how the levels of these two internal conditions affect the three answering behaviours and will discuss later their possible origins.

General arousal or drive affects all three answering behaviours. One is tempted to speculate, in accordance with Epstein's theory, that depending on its level, general arousal will either enhance or constrict awareness, and this, in turn, might well affect the student's ability to perceive subtle word meanings and relationships during his reading of the question. It might also affect the amount of memory which can be scanned looking for the correct answer. In addition, general arousal, or drive, will affect the ability of the student to choose the correct alternative from among those offered. Spence and Spence, in the discussion of their theory, showed that arousal level affected the student's ability to choose alternatives of lower habit strength and hypothesized that it was because of the multiplicative relationship between arousal and excitation. Therefore, one might expect that the effect of general arousal on the ability to choose the correct alternative will, to a large extent, vary from item to item and depend on the internal structure of the question. Note that in our discussion, in the first case, that of awareness, arousal level has an affect consistent with the inverted U shaped relationship, whereas in the second case the relationship between arousal and behaviour is monotonic and multiplicative.

The inappropriate focusing behaviour level is the tendency of the
student to focus on his feelings, to channel his energy towards unproductive self-deprecatory ruminations and other such behaviours. These are the behaviours and activities surmised by Wine in her work. At any point where the student is engaged in them they will adversely affect all three of the answering behaviours, although probably reading will suffer most, as the most likely effect here is that he will not attend to certain key words and will not be aware of the fact that he missed them. Also, on a timed test he will have wasted some of the allotted time.

The magnitude of each of these levels is determined by a number of factors. Each of these factors is complicated in itself, and they often interact with each other to intensify their results. In this section an attempt will be made to outline some the major factors operating, and then to suggest some of their possible origins.

General arousal level has three overall determinants, base level, situational level, and individual level. Base level can be visualized as the person's average anxiety level (that which just getting up in the morning will cause). The origins of this anxiety are probably those suggested by the psychoanalytic and cognitive theorists, notably Rogers, Freud, and May. Unfortunately, this is the kind of anxiety least amenable to simple reduction within the testing situation. On the other hand, we might hypothesize that it has a minimal influence on performance, as the individual is used to this level of anxiety and is used to functioning with it. Of course, it will have some effect, and this is likely to be part of the correlation observed between the Taylor MAS and test anxiety (which will be detailed later in the section on test anxiety measures).

The situational level is determined by a number of facets of the test.
The difficulty of the material and the normal difficulty of the tests of this type are two. The lighting, the crowdedness, the atmosphere, the importance, the instructions, and features like these will also affect this level. Moreover, the subject's estimate of how well he knows the material will be a factor. These factors, for the most part, will be the ones affected by the manipulations of this experiment.

The individual level has a number of major components. First is the reaction potential of the person. Given that the individual is exposed to a threatening stimulus of known magnitude, he will react with a rise in arousal that is unique to him. This reaction potential will then interact (probably multiplicatively) with the situational components and with the other individual components to produce an arousal level.

The other individual components are more wide ranging. Most important will be the extent to which the individual tends to view any sort of test as a threatening stimulus. The origin of this tendency is most probably a combination of those hypothesized by the conditioning theorists and the Freudians. First, what seems to be a phobia of tests is conditioned by the subject being punished after having taken some sort of test. Assumably, being punished after having done some sort of evaluative task is not an uncommon occurence and the phobia will then generalize to any task having the same characteristics, which are probably bound up in its being evaluative in nature. In line with the Freudians, one might suppose that the punishment in these cases was the withdrawal of parental love, that is, part of the classic separation of mother and child. This facet of test anxiety, as will be detailed later, is best treated using behaviour modification techniques that mitigate the effects of early conditioning.
Another individual factor, one which interacts with situational factors, is the extent to which the individual will underestimate his likely score. As was mentioned above, test anxiety correlates negatively with optimism. Therefore, one might expect that, given a standard expectation of success, the test anxious person will actually expect a lower score.

Note how these factors interact. For example, a student comes to a test with some likelihood of success, depending on his knowledge and ability. If he is test anxious, his expectation will be lower than his 'real' likelihood. This worry caused by low expectancy (quite rational aside from the estimation error) then adds to the worry engendered by the importance of the exam. This total is accentuated by the situational factors, the conditions in the exam room, to give a measure of perceived threat. Then this perceived threat is multiplied by the individual's reaction potential and this total is added to his base arousal to give a magnitude to his general arousal level.

The innappropriate focusing behaviour level is more difficult to explain. Because it is actually behaviours it seems reasonable to hypothesize that these behaviours are learned. However, none of the theorists reviewed here have come up with reasonable situations in which these behaviours might have been learned, and so, for the moment, we must accept their origin as unknown.

As was stated above, the likelihood of a correct response varies from item to item in a test. As we have seen, the difficulty and the habit strength of alternatives will affect the ability of the student to choose a correct response. Note that this is an immediate result of each particular item on the student's likelihood to be correct on that item.
More interesting is the effect of the item before on the student's likelihood of success on the present item. Terry and Isaacson (1971) did a study in which they inserted four questions with no possible correct answers into a test. By varying the order of the other questions they were able to ascertain that students did almost twice as badly on "post-impossibles" as they did on items preceding the impossible questions. They then reexamined their data and found that the differences were even more striking ($p = .01$) for students high in test anxiety. The exact mechanism of this effect is uncertain although one might imagine that one component is that the high anxious student will suppose that he is doing much worse than is true because he has missed one item.

Later it will be shown that there is reason to believe that students suffer more test anxiety on longer tests than on shorter ones and that some effects of test manipulation only show up on the second half of tests. If this supposition is true, it will change the likelihood of correct responding with time. It should be noted that this effect, if true, is contrary to Epstein's expectation of habituation. There are no published speculations which might account for its origin.

In summary, it must be said that if the conjectures presented above are true, that is, if test anxiety effects are best described using a systems model, one might expect confirmation by prediction. Unfortunately, essential ingredients in this model are the individual factors. Because of their complexity determining their magnitude is virtually impossible and without such a determination, prediction is impossible. Therefore, any confirmation of the system model will have to await improvements in the technology of assessment.
Test Anxiety Alleviation Techniques

There are a number of ways of reducing the effects of test anxiety that are generally in use today. Although some of these involve changes in the testing situation, most focus on the severely affected individual and are essentially psychotherapeutic techniques designed to alleviate, for the individual, the effects of the anxiety on his performance. The most effective of these come under the general heading of behaviour therapy. Ones that report success run the full gamut of these types of techniques, from simple systematic desensitization (Wolpe, 1958) through accelerated massed group desensitization (Richardson and Suin, 1973) and implosive therapy (Dawley and Wernich, 1973), to covert reinforcement of imaginal appropriate behaviour (Wisocki, 1973).

The only serious variations on these techniques is their combination with other ideas such as study counselling. In one such study (Allen, 1973) these combinations were shown to be more effective than the simple treatments themselves, but only marginally so.

Unfortunately, because of the expense, time, and stigma involved, these techniques are useful only in the most extreme circumstances, where the individuals involved are handicapped to a severe extent by their test anxiety. For the vast majority of affected individuals, where losses in performance are on the order of 10%, these techniques are irrelevant. What would be useful for these people is a simple, straight-forward method of alleviating test anxiety requiring neither large commitment of resources nor particular attention on the part of the student. It would seem that these would then necessarily be manipulations of the testing situation itself. There have been some efforts in this direction and these are
detailed below.

The most commonly used manipulation to reduce anxiety in test situations is one often used when I. Q. testing is done on children. The children are told that the I. Q. test is, in fact, a 'game' and that they should just relax, try to do their best, and have fun (Young and Brown, 1973). Under these conditions test anxiety has little or no effect. These results are consistent with a large amount of research that tends to show that if a test is not evaluative or ego-involving, test anxiety will not be present. This effect can also be seen in adults. S. Sarason (1957) found that when he told subjects that the verbal learning task they were doing was an I. Q. test they did significantly worse than when he told them that they were just doing the experimenter a favour by doing the experiment. I. Sarason and Palola (1960) showed a similar relationship when comparing neutral instructions with ones emphasizing that the digit symbol task they were using was an I.Q. test. And Long and Bessemer (1971) found that they could elicit the same response, although in their experiment they found that they had to mention three facts: (1) that the experimental verbal learning task was a test; (2) that it was an I.Q. measure; and (3) that it might be evaluated against the results achieved by the subjects' peer group.

All these experiments imply that if one could remove the evaluative component from the testing situation, test anxiety would have only the most minimal impact. Unfortunately, when one is dealing with adults, and when they know that their results are meaningful, as they must when they are taking normal college tests, there is no way to remove the evaluative component from the testing situation, giving this approach virtually no practical utility in college testing situations.
Another commonly used technique that might incidentally reduce the effects of test anxiety is to give frequent tests. Both Wallin (1975) and Dustin (1971) found that giving frequent tests in a psychology course almost completely neutralized the effects of test anxiety. Unfortunately, Marso (1970), who also studied frequent testing, found no such neutralizing effect. It seems likely that this difference might stem from the fact that Marso's tests were somewhat longer, being twenty-eight items long, compared to Dustin's ten items and Wallin's average of nineteen items. If this is true, then it might indicate that it is really the length of the test that caused the effect. This possibility fits well with data, to be presented below, that indicates that test anxiety has more of an effect in the latter parts of a test than it does at the beginning of a test.

A situation that seems to combine all of these effects, that is, frequent testing, short tests, and lessened evaluative pressure, is that which prevails in PSI or Keller method courses. One study focusing on test anxiety (Allen, Giat, and Cherney, 1974) was particularly noteworthy because both trait and state anxiety were examined. Trait anxiety had no correlational relationship with achievement, and state anxiety was shown to reduce in magnitude with each new test.

Whether a test is 'timed' or 'untimed' seems to make a difference in the degree to which it elicits test anxiety. Siegman (1956), Mattarazzo, Ulett, Guze, and Saslow (1954), and Morris and Liebert (1969) all reported this effect, even when, through yoking, the amount of time used in both conditions was the same.

Changing the stimulus qualities of the test itself can have quite dramatic effects. Smith, Ascough, Ettinger, and Nelson (1971) had real
success by changing 10 of 30 items on a normal psychology course exam from normal questions to humorous though valid ones. An example of a normal or non-humorous item is:

Over the past six years, Tom's behaviour has become increasingly more disturbed. He has developed a delusion that somebody is controlling his mind and he is also having bizarre visual and auditory hallucinations. Which other member of Tom's family is most likely to exhibit bizarre behaviour?
- choices: a) his mother b) his father c) his sister d) none of the above

This item in humorous form was:

Claiming to be a slot machine, Julius has been standing against a wall in a Las Vegas casino for six years making bell-like sounds and occasionally complaining that he is being tilted. Which other member of Julius' family is most likely to exhibit bizarre behaviour?
- same choices.

The average difficulty of these humorous items was found to be the same as their corresponding non-humorous items and they probed the same basic information. Still, their presence erased the debilitating effects of anxiety for the highly test anxious in the class while not affecting the scores of the low anxiety group at all.

One unfortunate aspect of all the methods listed above is that they are not very generalizable. One can imagine many situations in which any or all of these interventions might not be practicable. For this reason, this experimenter has chosen to concentrate on other, more generalizable methods. These will be described below.

The first and simplest of these is one developed by McKeachie, Spollie, and Speisman (1955) and labelled 'abreaction' by Bucky (1972). In this technique room is left on the answer sheet for students to write
comments. A direction is given for the students to: "Feel free to comment in the space provided." With this instruction McKeachie and his colleagues got significant rises in exam performance. Subsequently, they tried a second experiment in which they varied the directions given to the students. They found that with more specific directions, such as: "tell how you feel about the item" (which focused on the students' feelings), or "state explanations of answers when necessary" (which focused on student clarification), or "state your feelings and give an explanation" which focused on both), the facilitative effect disappeared. McKeachie et al. hypothesized that with more specific directions the student, rather than getting less anxious, actually became more anxious because he now worried that he might not be performing up to snuff on his new task, that of giving comments.

It is, of course, difficult at this juncture to say just why this effect occurs. Probably it is best to say rather grossly that commenting gives the anxious student a chance to work off his tensions. This offhand explanation is given some support by research done by Malmo (1966) in which he showed that anxious people differed from non-anxious people in that they recover more slowly from being startled. As a measure he chose striate muscle activity. In non-anxious people and in anxious people stimulation designed to startle causes a mean rise in muscle tension. In non-anxious people this rise in tension reduces to the prestimulus level in less than a second, whereas in anxious people mean muscle tension level remained elevated significantly higher than that of non-anxious people for a period longer than three seconds. So, it does not seem unreasonable to ascribe the beneficial effects of commenting to a release in tension in the anxious person.
It should be noted that McKeachie et al. merely hypothesized a connection between commenting and anxiety. They did not actually establish it experimentally. It remained for Calvin, McGuigan, and Sullivan (1957) to perform the necessary experiment to tie commenting with anxiety. In their experiment they classified subjects by their score on the Taylor Manifest Anxiety Scale and found, as they expected, that it was people who were high in manifest anxiety who benefitted from the commenting condition.

One interesting aspect of the commenting condition was that both McKeachie et al. and Calvin et al. found that the facilitative effect of the commenting was much more pronounced in the second half of the test than in the first half. Bucky (1972) examined this effect in much more detail. He gave the California Test of Basic Skills to a group of thirteen to fifteen year old children, along with extra pieces of paper on which they were to put their comments. This particular test has a number of subsections, and Bucky found that the facilitative effect of commenting increased with each subsection, and that the degree of facilitation for those students high in test anxiety correlated with the amount of commenting that they did.

It is interesting to note that this cumulative effect of test anxiety has been seen in a number of other experiments. It has been shown that if easy items come first in a test, then a highly anxious student will do better than if hard items come first (I. Sarason and Palola, 1960; Hambleton and Traub, 1974). Similarly, Long and Bassemer (1971) showed that if a student has been doing well in the first part of an experimental task, then ego-involving instructions administered in the middle will
have no effect. However, if the student has been doing poorly, and he is highly test anxious, then those same ego-involving instructions will severely affect his performance.

It might follow logically from these experiments that one way to reduce the effects of test anxiety would be to interrupt the tests themselves with a break so that students would have a chance to relax and lose some of their tensions. This seems, at least in the simple case, not to be so. Morris and Perey (1972) experimented with interruptions, both relevant and irrelevant. They found that the only affect they could initiate was a rise in hostility among those students who had been irrelevantly interrupted. They found no changes in test anxiety or in emotionality and also, as an aside, found no main effects at all for interruption on achievement. It should be noted, though, that all their interruptions were vaguely unpleasant ones, as they were focusing more on stimulating rises in anxiety. It might not follow from their study that an interruption designed to calm the test anxious student would not have an advantageous effect. In fact, it could by hypothesized that it is just such a calming interruption which explains the efficacy of one of the behaviour modification methods that relies on the use of 'cues' by the student. In a study done by Russell and Sipich (1973) the patient was taught to say the word 'calm' to herself when anxiety threatened and then was taught to pair this word with relaxation. Perhaps more overt calming methods would work for the general, untrained, student population.

Another effective method for reducing the effects of test anxiety that is easily generalizable, is the playing of music either during or before a test (Stanton, 1973 and 1975, respectively). In his original
research Stanton played soft, non-distracting, classical music to experimental groups of primary, secondary, and tertiary (college) students taking one of their normal exams. In each case a matched control group was given the same exam in the normal silent way. He found no significant differences for either of the first two age groups, but did find that in the college group students high in test anxiety performed significantly better in the exam room where the music was playing than a matched group of test anxious people did in the silent room, their group means being 70 and 54 respectively.

In a second experiment in this first study, Stanton examined the effect of music on a laboratory task, the memorizing of the order in which geometric shapes were presented. Again he found no effect on students of primary school age, but did find a significant interaction of music with test anxiety for both the secondary school students and the college students.

One anecdotal result of the study was that students soon became accustomed to the music in the exam room and stopped noticing it. In a second study Stanton (1975) examined this effect more closely. He tried two different interventions, the first, playing music throughout an experimental task, as in the first study, and the second, playing music only as background while the students were entering the experimental room and while they were listening to the experimental instructions. He found, as can be seen below, that both of these interventions were equally good at alleviating performance decrements caused by test anxiety. (Means for high and low anxious groups, respectively, were: in silence- 12.2 and 14.0; with music played only at the start - 14.7 and 13.6; and with music played throughout - 14.5 and 13.7.)
One problem that has been suggested with playing music is that it might interfere with the concentration of some student, and would lead to many complaints by such students, particularly those not affected by test anxiety. Fortunately, Stanton found that more than two-thirds of the students who experienced music stated a preference for it, and that very few students actually stated that they found the music distracting. Furthermore, of the few who did complain about the music, Stanton found that a number of them actually recorded perfect scores even though they said that they were distracted.

While commenting and music have been tried in real exam situations, there is a third method, tried only as yet in the laboratory, which might be easily generalizable to general exam situations. This third method is a logical outgrowth of a method, developed by I. Sarason, Pederson, and Nyman (1968), for treating highly anxious individuals. They would show their patient a model behaving appropriately in a testing situation, and demonstrated that this tended to mitigate his anxiety.

In the extension of this method, I. Sarason (1975) focused on the type of model that he was providing his test anxious students. In his experiment, Sarason attempted to influence success at a nonsense syllable task with three different styles of model. In each condition the subject would enter the experimental room and be greeted by the 'assistant' running the experiment. This assistant would chat with the subject for a moment or two while setting up the memory drum. During the conversation, one of four topics would be discussed. In what he called the 'high-anxious coping' role the assistant would state that she got nervous and performed poorly on tests, except that she had resolved this problem by remembering four
methods of coping with test anxiety. (These methods were:

1. reminding yourself periodically to stop thinking about yourself and to concentrate on the task at hand;
2. thinking about aspects of the task that might be especially interesting to you;
3. not allowing yourself to get flustered by errors and difficult items, but to keep working on the task at a steady pace; and
4. forcing yourself not to think about other people and how they will or might perform the task.)

In the high-anxious, non-coping, role, the assistant would describe how she got nervous and performed poorly on tests, but would not describe any successful coping mechanisms. In the 'low-anxious' role, the assistant would talk about how calm and self-assured she felt trying tests. And, of course, there was a control and a placebo group. In the placebo group the assistant talked about campus life in a friendly manner, and in the control group the assistant remained essentially silent while setting up the equipment.

Sarason's original hypothesis was that it would be the 'disclosing' aspect of these models that would provide the most facilitation. However, his data showed that mere disclosure actually worsened performance. Only in the disclosure with coping group was facilitations seen. (See Table 1.)

It would seem that this technique would be easily transferable to a normal exam situation. With almost no real deception, almost any exam proctor could describe his or her own feelings of anxiousness and then go on to describe the methods listed above for dealing with it. And, it is entirely possible that the real manipulation in this experiment was merely
the assistant telling the student the anxiety reduction techniques, thereby legitimizing any anxious feelings the student might have, and also giving the student a real means of coping with those feelings. Of course, whether this technique will, in fact, work with a large group, must be tested empirically.

As has been mentioned, the first two of these anxiety reduction techniques have already been proven in real-life exam situations. Therefore, in addition to being a replication of these studies, this study will attempt to cover new ground. There is some legitimate question as to whether the effects achieved with music and commenting are 'real' effects, or whether they are 'Hawthorne' effects, that is, effects whose cause is merely the existence of an experiment. In this study, the conditions will be present for two exams, the assumption being that any 'Hawthorne' or other transient effect will not show up on the second exam in the course, whereas any real effect of commenting or music will.

Test Anxiety Scales

There are a wide variety of scales available for measuring test anxiety, and some comment on them is necessary so as to understand the choice of instrument made here.

The first studies of the effects of anxiety on test performance used the Taylor Manifest Anxiety Scale (MAS) (Taylor, 1953) as a measure of test anxiety (for example, I. Sarason, 1957). However, as has been mentioned above, there is a limited correspondence between test anxiety and general anxiety and this instrument was soon superseded by instruments designed especially for measuring test anxiety.
The Test Anxiety Questionnaire (TAQ) (Mandler and S. Sarason, 1954) was the first of these and, in its various forms, is still the most popular. It has been adapted for high school students (and called the Test Anxiety Scale - TAS) and for children (the Test Anxiety Scale for Children - TAS-C). It has been mentioned that this scale was factor analyzed by Sassenrath (1964) and Gorusch (1966) in their attempts to discover underlying factors in test anxiety. It is interesting to note that Gorusch identified a single third order factor during his analysis, and surmised that this must be the test anxiousness or main factor probed by the scale. According to his analysis, this third factor was sufficiently weak so as to indicate that the scale itself should be reconstructed.

In 1958 I. Sarason developed a variation of the TAQ which he called the Test Anxiety Scale (TAS). (It should be noted here that having two different TAS's by two distinct Sarasons created havoc in the literature). Sarason originally favoured use of his scale because it had 'true-false' items and this, he said, made it easier for use by psychiatric patients. Still, since I. Sarason is one of the premier researchers in this field, his TAS has been used extensively with normal groups of subjects both by himself and by other researchers. His TAS correlates with S. Sarason's TAQ with an average correlation of .94.

Following the Sarasons, Alpert and Haber (1960) developed the Anxiety Achievement Test (AAT), the one that will be used in this study. This test has two subsections intended to tap two factors that they hypothesize exist in test anxiety, namely, facilitating anxiety and debilitating anxiety. Debilitating anxiety is the kind that has been the focus of the discussion here so far. Facilitating anxiety is hypothesized to be emotional
arousal that leads to task relevant behaviour and thus to increased achievement. It is interesting to note that most researchers choose to ignore these separate scales and instead derive a 'difference' or combined score from the AAT. They say that this combined score is a better predictor of achievement than either scale taken alone.

There are two reasons for preferring the Alpert-Haber scale. First, having both subscales available allows one to make more interesting interpretations of the final data. The second reason is that Alpert and Haber have reported data that shows that their test is superior to both Taylor's MAS and S. Sarason's TAQ at predicting GPA, course grade, and final exam grade. They reported that high performance on these correlated, on average, -.28 with high anxiety on their scale compared to -.23 with the TAQ and -.13 with the MAS. Carrier and Jewell (1966) did a similar study in which they showed that the AAT was superior to I. Sarason's TAS and to the Anxiety Differential (to be discussed below). In their study the correlations of the AAT with achievement on an exam ranged from -.27 to -.50.

It should be mentioned that there is dispute in the field as to whether a 'trait' anxiety scale, such as the ones listed above, or a 'state' anxiety scale is better. There are two major 'state' anxiety scales. The first is the State-Trait Anxiety Inventory (STAI) developed by Spielburger and Gorusch (1966). The second is the Anxiety Differential (AD) developed by Alexander and Husek (1963). It is a semantic differential instrument originally developed by Alexander and Husek to measure anxious responses prompted by viewing films of car accidents, medical operations, and mutilated bodies. It was modified slightly for use as a state anxiety measure for testing situations.
Tennyson and Boutwell (1973) showed that a within-task state measure was better than a pre-task state measure or a trait measure at predicting performance on a concept acquisition task.

On the other hand, Hedl (1972), using regression techniques, showed that state anxiety outside of the testing situation was just as good as state anxiety within the testing situation at predicting TAS and that trait anxiety was much better. Of course, there is nothing very surprising about the fact that one trait measure would be better than a state measure at predicting another trait measure. Unfortunately, Hedl neglected to correlate performance on his task with either his trait or state measure of anxiety. He did confirm, however, the fact that state anxiety score is changed dramatically by presence at an exam.

One reason that one might want to correlate performance with a personality variable such as test anxiety, is that one might want to prescribe different treatments for different types of people. For example, if one were to find out that highly anxious subjects performed best in one condition, and low anxious subjects performed best in another, one would be tempted to divide the class before each exam, and to place students in the condition where they would be expected to have the most advantage. Allen (1970) tries to make the argument that state anxiety would be a more appropriate measure with which to examine such 'aptitude treatment interactions' because it seemed logical to him that a state anxiety measure would be more valid. This experimenter disagreed with that position for two reasons. First, score on a within-test state anxiety measure would be too dependent on the situation associated with one particular test. For instance, if the student did not see that particular
test as being ego-involving, then his test anxiety would not emerge. In that case, one could not be sure that the student would get the same anxiety score every time, which means that the reliability of the measure would be in doubt. And, validity can only be as good as reliability, at its best. Secondly, it would be difficult to simulate real exam pressure and consequent anxiety early in the semester when one would want to make any test prescriptions. For these reasons, it seemed that a trait measure would be appropriate for use in this study.
Pilot Study: Description, Results, and Discussion

During the fall semester of 1975 a pilot of the experiment was run, using students in a child psychology course as subjects. The students were approached just before their second course exam and were randomly assigned to groups which met to take their exams in specific rooms. (If a student did not wish to participate in the experiment, he was excused. Less than 10% of the students chose not to participate in the experiment.)

There were seven groups: one control group and two groups each in the music, coping model, and commenting conditions. These groups’ conditions were as follows:

Control - In this condition students took their exam exactly as they normally would in a course of this type.

Music - In this condition soft music was played in the background throughout the exam.

Coping Model - In this condition the exam proctor shared with the students the revelation that he too suffered from test anxiety, and that he had successfully overcome it by heeding a few simple hints. Then a sheet with the hints on it was handed out.

Commenting - In this condition students were allowed to make comments about the exam on a separate sheet of paper.

At the end of their exam, the students were asked to complete the debilitating scale of the AAT. Complete data was obtained from all students.

A correlational analysis of the data was performed. It was found that correlations between performance on the exam and high test anxiety were: Control : +.04
Music : -.25.
Coping : -.17
Commenting: -.06.
A similar correlation was made between high test anxiety and student performance on the first exam in the course, an exam that had been taken before the experiment was started. These correlations were:

- Control group people: +.27
- Music group people: -.35
- Coping model group: -.06
- Commenting group: +.02

Now, if the results were as expected, we would have found that the control group in both conditions and the other three groups in the first exam (the second list of correlations), would show a correlation between high anxiety and performance of -.30 or thereabouts. We would then expect that the experimental conditions would reduce this correlation in the second exam to a correlation of about -.10, allowing one to conclude that the conditions had alleviated some of the effects of test anxiety.

Unfortunately, an examination of the data shows unexpected effects. Probably the most unexpected is that of the correlations shown by the control group on both exams and the other groups on the first exam. With only one exception (the music group on the first exam) the expected negative correlation of high anxiety with performance failed to appear. Therefore, it becomes impossible to ascertain whether or not the experimental conditions had any effect, even though their correlations fell within an acceptable range of the expected correlation of -.10.

There are a number of explanations possible for these surprising results. First, because the test anxiety scale was administered after the second exam, it is quite possible that there was some relationship between a student's performance on the second exam and his test anxiety score. If this were true, then one would expect a spuriously high correlation
of the test anxiety score with score on the second exam, and a spuriously low correlation of the test anxiety score with score on the first exam.
(Note, these correlations would be negative.)

A second explanation can be found in the difficulty, or lack of it, found in the first exam. The mean score of the class on the first exam was 86%. As has been discussed, difficulty on an exam heightens the effects of test anxiety. Therefore, it would stand to reason that simplicity would minimize these effects, and this might very well explain the small effects of test anxiety on scores on the first exam.

The mean score on the second exam was 75%. Considering the high mean score on the first exam, one can only suppose that the contrast between the two exams heightened the effects of test anxiety on the second exam, biasing the data in the wrong direction.

Two simple changes can be suggested to correct these defects. First, the anxiety scale should be given early in the semester so that it yields a score more valid for predictive purposes. Second, a class should be chosen where one would normally find test anxiety effects, that is, where the exam can be expected to be difficult and threatening enough to produce test anxiety effects in the class. Both these changes are easily affected.
Rationale of the Present Study

The main experiment in this study was designed to answer two central questions. First, it was designed to discover whether certain specific examination conditions would serve to mitigate the debilitating effects of test anxiety on performance in a real college testing situation. Second, it was designed to discover whether these conditions could continue to provide their beneficial effects when used repeatedly. That is, it was designed to discover whether any effectiveness these conditions had was a transient phenomenon.

In order to answer these questions it was decided to conduct this experiment on two separate exams in a single course. If the conditions employed had the expected beneficial effects on the first exam it would be possible to administer the same conditions during the second exam and to discover whether their beneficial effects recurred. If the conditions were not successful on the first exam, it would be possible to reapply them on the second exam in hopes of replicating earlier research which indicated that they would, in fact, prove beneficial.

While this main experiment was in progress, the opportunity presented itself for the experimenter to conduct two supplementary experiments dealing with questions related to the ones listed above. In the first supplementary experiment, the experimenter was allowed to solicit volunteers from a course who would allow themselves to be randomly assigned to various experimental conditions. It was hoped that this experiment might yield data to corroborate any findings of the main experiment, and further might yield interesting data on the nature of students who would choose to participate in this sort of experiment.
In the second supplementary experiment, the experimenter was allowed to solicit subjects in a course provided these subjects were told exactly what the experimental conditions would be. It was hoped that data from this experiment would also corroborate findings from the main experiment. Moreover, it was possible in this experiment to examine a number of different experimental conditions and to discover whether students would, if given the choice, choose to endure examination conditions noticeably different than the norm.
Method

Subjects

The subjects in this experiment were 245 students in an introductory psychology course with an enrollment of 285. (Of these 40 students who were not subjects in the experiment, 30 were not because they took their examination at other than the scheduled times, and 10 were not included because they chose not to be.) Of the 245 students, 20 did not take the second scheduled exam and so data for these subjects is incomplete.

Measures and Materials

The performance measures were the first two of three scheduled examinations in the course. Each was a 50 item, multiple-choice exam which examined the students on material covered in that third of the course.

The test anxiety scale was the Alpert-Haber Achievement Anxiety Test. (See Appendix 1.) This scale allows one to identify a score for debilitating anxiety and a score for facilitating anxiety. (In this study a high score on the debilitating anxiety scale indicates that the student is high in debilitating anxiety. A high score on the facilitating anxiety scale indicates that the student is high on facilitating anxiety.) These scores can be combined to yield a combined or total test anxiety score. (This is done by subtracting the facilitating anxiety score from the debilitating anxiety score and adding a constant, in this case 50. That is, "combined = deb - fac + 50". This formula yields a score on the combined anxiety which is high when the student
is highly test anxious.)

A questionnaire was given to the students after their second exam. This questionnaire (Appendix 2) was developed by the experimenter.

Each student in the commenting condition was given a 'comment sheet'. This sheet was designed by the experimenter. (See Appendix 3.) Note that this sheet was printed in blue ink, whereas the examinations were printed in black ink. This identified the comment sheet as being associated with the experiment rather than identifying the comment sheet as being a regular handout of the course instructor's.

Each student in the hints condition was given a sheet of hints designed to help a student allay any feelings of test anxiety normally affecting that student. This sheet (Appendix 4) was designed by the experimenter.

The background music used in this experiment was recorded on a tape cassette which allowed for one-half hour of music on each side. One side, one-half hour's worth, was of Mozart flute quartets while the other side was a half hour of the quieter sections of Rimsky-Korsakov's Scherezade. The music was played on a monophonic tape player into two large speakers positioned in corners of the experimental room.
Procedure

On the first day of the semester the experimenter went to the course lecture hall and explained the goals and design of the experiment to the students in the class. The students were told that the experimenter hoped to discover ways of reducing the effects of test anxiety by making simple changes in the environments of exams. They were not told what the exact experimental conditions would be.

The students were told that they would be divided randomly into groups and that each group would go to a different examination room to take its first two exams in the course. They were told that each room would have a different environmental change and that these changes might be subtle enough that they would not notice them.

Students were told that, because this was an experiment that was being done on real exams, they were not obliged to participate. If they did not wish to be in the experiment, they were told to ask the instructor or his assistant where they should go on the exam date to take the exam in a room where there would be no manipulation of the environment. Because only ten students made such a request, they were sent to the room where the control group was taking its exam. Note that data was not collected from these ten students.

Just before all these explanations were made on the first day of the semester, the anxiety scale was administered to all students present that day. Students were told that their scores would be kept totally confidential and were asked only to identify themselves by student number. It was explained to the students that they were not obliged to fill out and hand in the anxiety scale if they didn't want to. As far as could be
seen, no student failed to complete and hand in an anxiety scale.

For the experiment students were randomly assigned (on the basis of the last two numbers of their student identification number) to one of six groups. They were informed of their assignment two weeks before the exam both in a written handout and verbally. The handout was available and the verbal announcement was repeated at every class period until the exam.

On the day of the exams each student went to his or her assigned room to take the exam. (There were a few students who went to the wrong rooms. No attempt was made to correct these errors as it was assumed that it would not introduce any systematic differences between the groups into the experiment.) Each room had at least one graduate student proctor present, one who was knowledgeable in the course material. All students in each room began the exam at the same time and left whenever they were finished. There was no time limit on either exam. The exams were closed-book, multiple-choice exams and were taken by all students in standard fashion, except for what differences might be caused by the experiment conditions. The conditions are described in the section below.

Following the second exam each student was given the questionnaire to fill out indicating their level of anxiety and their reaction to the experimental conditions.
Conditions

Commenting

Two groups were assigned the commenting condition for both exams. In this condition each student was given a special 'comment sheet' (Appendix 3) and was told before the exam began: "Feel free to comment about the exam on the special comment sheet provided."

Music

Two groups were assigned the music condition for both exams. In this condition soft background music was playing in the exam room when the students arrived. The volume of the music was lowered a little when the exam began and continued to play until all students had completed their work.

Hints or Coping Model

One group was assigned to the coping model condition for the first exam. In this condition the proctor attempted to convey to the students that he too suffered from test anxiety and that he had managed to overcome his problem with the help of a number of hints. His speech, delivered extemporaneously, approximated the following:

"Now, as you know, the conditions in the other exam rooms are intended to reduce the effects of test anxiety in those who are normally troubled by it. I'm sure you aren't surprised to know that anxiety is a very common experience in exams. I myself feel it fairly often. Still, I've learned to cope with it quite well just by remembering a few simple exam-taking hints. We thought it might be interesting to see if these hints help you as much as they've helped me so we've prepared a handout listing them for each of you."
Once this introduction was completed, the list of hints (Appendix 4) was handed out and the proctor read the list aloud to make sure that all students could read them. Immediately following this reading the exams were handed out and the testing session proceed normally.

**Combination Condition**

For the second exam, the group which had been in the coping model condition for the first exam was exposed to all three of the experimental conditions at once. In this condition (labelled "all" on the tables in the results section) the students were greeted by music when they arrived at the exam room and this music continued playing, just as in the music condition, throughout the exam. Before the exam began each student was given a comment sheet with its accompanying verbal instruction, and each student was given a list of the exam-taking hints along with the speech outlined in the section immediately above. An extra proctor was assigned to this group for the start of the exam and consequently the added time taken to hand out the extra sheets was negligible.

**Control Condition**

One group was assigned to the control condition for both exams. In this condition students came into the room, were given their exams when the time to begin them arrived, and left whenever the finished, except that after the second exam they were asked to fill out the anxiety self-report measure. The proctor for the control condition was the course instructor. The students who attended the control condition because they did not want to participate in the experiment were interspersed among the other students quite randomly and, as far as could be seen, did not tell the other students why they came to be in that particular group.
Method - Supplementary Experiments

First Supplementary Experiment

Subjects

The subjects in this experiment were students in an abnormal psychology course. Forty subjects volunteered to expose themselves to the experimental conditions and another 157 students completed the test anxiety scale. There were approximately 100 other students in the course. These students were absent on the day the anxiety scale was administered and consequently could not be included in the experiment.

Measures and Materials

The performance measures were the first two of three scheduled examinations in the course.

The test anxiety scale (Appendix 1), the 'comment sheet' (Appendix 3), and the sheet of hints (Appendix 4) were used in this experiment and were identical with the ones used in the previous experiment.

Procedure

During the third week of the semester the experimenter went to one of the course lectures and explained the goals and design of the experiment to the students in the class. The students were told that the experimenter hoped to discover ways of reducing the effects of test anxiety by making simple changes in the environment of examination rooms. They were not told what the exact experimental conditions would be but they were assured that the conditions would be very benign and not distracting at all.

Just before this explanation was made, the anxiety scale was adminis-
tered to all students present that day. As in the previous experiment students were told that their scores would be kept totally confidential and were asked only to identify themselves by student number. As in the previous experiment, it seemed as though all students present completed and handed in their anxiety scales even though they were told that they were under no obligation to do so.

Approximately three weeks later and ten days before the first exam, the experimenter returned to the class during a lecture period to solicit volunteers for three experimental groups. The students were not told what the experimental conditions would be but were again assured that they were benign and not distracting. Students volunteered by signing their names to one of three lists each of which indicated a different exam room for students to attend on the day of the first exam. Although students were not randomly assigned among the three rooms but rather chose them for themselves, there was no reason to believe that they would do so any way but randomly.

On the day of the exams each student volunteer went to his assigned room. All other students went to the class' normal lecture hall. Each exam room had at least one graduate student proctor present. All students in each room began the exam at the same time and left whenever they were finished. There was a time limit of one hour for the exam. All exams were taken in standard fashion except for differences dictated by the experimental conditions. These conditions were identical to the same conditions in the previous experiment. Note that in this experiment only the commenting and hints conditions were examined.
The first experimental group (arbitrarily numbered) was exposed to the commenting condition for the first exam and was exposed to the hints condition for the second exam. As stated, these conditions were identical to the same conditions described in the previous experiment.

The second experimental group was exposed to the hints condition for the first exam and was exposed to the commenting condition for the second exam. Note that the same proctor who had presented the hints to the first group during the second exam presented the hints to the second group during the first exam.

The third experimental group was exposed to the hints condition for the first exam. Because this group was quite small and could not be serviced easily, it was disbanded for the second exam and the students in it were told to take their exam in the main lecture hall.

All students who were not in these experimental groups took their examination in the lecture hall where the class normally met during the normal class period. These students were designated the control group. As was mentioned above, this group included the students from the third experimental group for the second exam.
Second Supplementary Experiment

Subjects

The subjects in this experiment were students in an adolescent psychology course. Thirty-nine students took their first exam in one of the experimental groups and another 150 students completed the test anxiety scale. There were approximately 100 other students in the course. These students were absent on the day the anxiety scale was administered and consequently could not be included in the experiment.

Measures and Materials

The performance measure was the first of three scheduled examinations in the course.

The test anxiety scale (Appendix 1) which was used in this experiment was the same as the one used in the previously described experiments.

The background music and the equipment on which it was played were identical with those used in the main experiment.

A thirty cup boiling water urn, styrofoam cups, instant coffee, tea, sugar, artificial creamer, spoons, chocolate chip and oreo cookies were all available in the 'coffee and cookies' condition.

Procedure

During the third week of the semester the experimenter went to one of the course lectures and explained the goals, design, and procedure of the experiment to the students in the class. The students were told that the experimenter hoped to discover ways of reducing the effects of test anxiety by making simple changes in the environment of examination rooms.
The students were told exactly what the experimental conditions would be and some effort was made to make the conditions sound attractive.

Just before the explanations were made, the anxiety scale was administered to all students present in the class that day. As in the previous experiments the students were told that their scores would be kept totally confidential and were asked only to identify themselves by student number. As before, it seemed as though all students present completed and handed in their anxiety scales even though they were told that they were under no obligation to do so.

Approximately three weeks later and ten days before the first exam, the experimenter returned to the class during a lecture period to solicit volunteers for three experimental groups. The students were told exactly what the experimental conditions would be. Students volunteered by signing their names to one of three lists circulated in the classroom. Each list had a brief description of one of the experimental conditions and a notice indicating in which exam room that condition would be in effect.

On the day of the exam each student volunteer went to the exam room where the condition he had chosen was in effect. Each exam room had at least one graduate student proctor present. All students in each room began the exam at the same time and left whenever they were finished. There was a time limit of one hour and fifteen minutes for the exam. All exams were taken in standard fashion except for differences dictated by the experimental conditions. These conditions are described below.
Conditions

Music Throughout the Exam

In one experimental group students arriving for the exam were greeted by soft background music. The volume of this music was lowered a little when the exam began but it did continue to play until all students had completed their exams and had left.

Music Before the Exam Only

In one experimental group students arriving for the exam were greeted by soft background music. The volume of this music was lowered somewhat during before-the-exam announcements (typographical errors on the exam and the like) and then was turned off completely when students began working on the examination.

Coffee and Cookies

In one experimental group students arriving for the exam were greeted with a table set up at the front of the exam room with boiling water, all the necessary fixings for coffee or tea, and a large plate of assorted cookies. Students allowed to fix themselves some coffee or tea and were allowed to take a couple of cookies, free. During the first half of the exam (until the water ran out), students were allowed to approach the table, one at a time, and fix themselves a refill.
Control Condition

All students who had not volunteered to be in one of the experimental groups took their exam in the lecture hall where the class normally met during the normal class period. These students were designated the control group, although themselves did not think of themselves as being participants in the study.
Results

Nine measures were collected from each subject in the experiment. These were: 1) the student's score on the Alpert-Haber Achievement Anxiety Test with both debilitating and facilitating scales combined by subtracting the facilitating score from the debilitating score and adding an arbitrary constant, in this case 50; 2) the student's score on the facilitating scale only; 3) the student's score on the debilitating scale only; 4) the student's score on the first examination in the course; 5) the student's score on the second exam; 6) the student's assessment of his anxiety on the second exam compared with the anxiety that student normally felt on an exam; 7) the students' assessment of the effects of the experimental condition on performance on the second exam, that is, whether the condition was helpful or whether it was distracting; 8) the student's subjective reaction to the experimental condition; and 9) the student's subjective assessment of the reaction of other students in the same experimental group. (Note that these last two measures were derived by the experimenter from open-ended responses by the students.

If the students had made no response, or if their response was judged to be neutral, the response was scored as a 3. If the response was somewhat positive it was scored as a 4 and if it was very positive it was scored as a 5. If the response was somewhat negative it was scored as a 2 and if it was very negative it was scored as a 1. The reliability of this scoring was checked by having a second graduate student rate 30 forms chosen from group 5, for measure 8, the student's subjective reaction to the experimental conditions. Of these 30 responses, 22 were rated identically, 3 were rated 1 point lower by the second graduate student, and 5
were rated 1 point higher. No response scores were more than 1 point different.) The means for all nine of these measures, listed both by group and for the whole class, are to be found in Table 1.

Because the subjects were randomly assigned to their respective groups it was expected that there would be no differences between groups on the anxiety measures. As a check a one way analysis of variance was performed for all three anxiety measures and, as expected, no significant differences were observed. (See Table 2.)

One way analyses of variance (by groups) were also performed on the scores each student received on both the first and the second examinations. Again, no statistically significant differences were observed between groups. (See Table 3.) This indicated that none of the experimental conditions influenced an entire group of subjects' performance level on either of the examinations.

The Relationship of Test Anxiety and Performance

In order to examine the relationships between scores on the test anxiety measures and performance on both examinations, two analyses were performed. First, a correlation was computed between performance on each exam and score on each of the anxiety measures, for the class as a whole. These correlations are listed in Table 4 and, as can be seen, the correlations between performance and scores on the combined anxiety and the debilitating anxiety scales are all significant, whereas the correlations between performance and facilitating anxiety do not reach statistical significance.

The relationship between test anxiety score and performance can be
Table 1.

Means of Experimental Variables by Group
(with standard deviations in parentheses)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Comment</th>
<th>Comment</th>
<th>Hints &amp; All</th>
<th>Music A</th>
<th>Music B</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A.</td>
<td>B.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n=35</td>
<td>=34</td>
<td></td>
<td>=50</td>
<td>=45</td>
<td>=40</td>
<td>=41</td>
</tr>
<tr>
<td>Total Anxiety</td>
<td>53.4</td>
<td>52.4</td>
<td>51.4</td>
<td>50.7</td>
<td>52.9</td>
<td>51.9</td>
<td>52.0</td>
</tr>
<tr>
<td>(hi is hi anxiety)</td>
<td>(6.6)</td>
<td>(9.3)</td>
<td>(7.9)</td>
<td>(9.7)</td>
<td>(10.2)</td>
<td>(8.4)</td>
<td>(8.7)</td>
</tr>
<tr>
<td>Facilitating Anx.</td>
<td>18.0</td>
<td>18.7</td>
<td>19.2</td>
<td>19.3</td>
<td>19.0</td>
<td>19.3</td>
<td>19.0</td>
</tr>
<tr>
<td>(hi is facilitating)</td>
<td>(3.7)</td>
<td>(4.8)</td>
<td>(4.9)</td>
<td>(5.1)</td>
<td>(4.6)</td>
<td>(4.0)</td>
<td>(4.6)</td>
</tr>
<tr>
<td>Debilitating Anx.</td>
<td>21.4</td>
<td>21.0</td>
<td>20.6</td>
<td>20.0</td>
<td>21.8</td>
<td>21.2</td>
<td>21.0</td>
</tr>
<tr>
<td>(hi is debilitating)</td>
<td>(5.5)</td>
<td>(6.8)</td>
<td>(5.5)</td>
<td>(6.3)</td>
<td>(7.2)</td>
<td>(5.1)</td>
<td>(6.2)</td>
</tr>
<tr>
<td>Score on First Examination</td>
<td>27.3</td>
<td>26.1</td>
<td>24.5</td>
<td>27.4</td>
<td>26.7</td>
<td>25.2</td>
<td>26.1</td>
</tr>
<tr>
<td></td>
<td>(5.7)</td>
<td>(5.8)</td>
<td>(5.5)</td>
<td>(6.3)</td>
<td>(6.0)</td>
<td>(6.5)</td>
<td>(6.0)</td>
</tr>
<tr>
<td>Score on Second Examination</td>
<td>25.8</td>
<td>26.1</td>
<td>25.4</td>
<td>27.9</td>
<td>26.7</td>
<td>26.7</td>
<td>26.4</td>
</tr>
<tr>
<td></td>
<td>(4.7)</td>
<td>(6.4)</td>
<td>(6.4)</td>
<td>(6.2)</td>
<td>(6.8)</td>
<td>(6.4)</td>
<td>(6.2)</td>
</tr>
<tr>
<td>Anxiety in Second Exam</td>
<td>3.20</td>
<td>2.83</td>
<td>3.05</td>
<td>3.03</td>
<td>2.93</td>
<td>2.97</td>
<td>3.01</td>
</tr>
<tr>
<td>(hi is low anxiety)</td>
<td>(.70)</td>
<td>(.83)</td>
<td>(1.04)</td>
<td>(.80)</td>
<td>(.84)</td>
<td>(1.02)</td>
<td>(.88)</td>
</tr>
<tr>
<td>Effect of Conditions</td>
<td>2.85</td>
<td>3.00</td>
<td>2.60</td>
<td>3.23</td>
<td>2.93</td>
<td>3.07</td>
<td>2.94</td>
</tr>
<tr>
<td>(lo is helpful, hi is distracting)</td>
<td>(.44)</td>
<td>(.37)</td>
<td>(1.00)</td>
<td>(.97)</td>
<td>(1.07)</td>
<td>(.25)</td>
<td>(.80)</td>
</tr>
<tr>
<td>Reaction to Condition</td>
<td>3.12</td>
<td>2.93</td>
<td>3.43</td>
<td>2.78</td>
<td>2.83</td>
<td>2.94</td>
<td>3.02</td>
</tr>
<tr>
<td>(hi is favourable)</td>
<td>(.42)</td>
<td>(.52)</td>
<td>(1.2)</td>
<td>(1.12)</td>
<td>(1.3)</td>
<td>(.36)</td>
<td>(.95)</td>
</tr>
<tr>
<td>Assess Others' Reaction to Cond.</td>
<td>3.12</td>
<td>2.90</td>
<td>3.08</td>
<td>2.93</td>
<td>3.00-</td>
<td>2.94</td>
<td>3.00</td>
</tr>
<tr>
<td>(hi is favourable)</td>
<td>(.33)</td>
<td>(.31)</td>
<td>(.83)</td>
<td>(.83)</td>
<td>(.80)</td>
<td>(.36)</td>
<td>(.64)</td>
</tr>
</tbody>
</table>
Table 2.

One Way Analyses of Variance of Anxiety Measures by Groups

### Analysis of Combined Anxiety Score by Group

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>5</td>
<td>193.83</td>
<td>38.77</td>
<td>.505</td>
<td>.777</td>
</tr>
<tr>
<td>Within Groups</td>
<td>239</td>
<td>18353.14</td>
<td>76.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>244</td>
<td>18546.96</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Analysis of Facilitating Anxiety Score by Group

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>5</td>
<td>53.08</td>
<td>10.62</td>
<td>.506</td>
<td>.771</td>
</tr>
<tr>
<td>Within Groups</td>
<td>239</td>
<td>5012.33</td>
<td>20.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>244</td>
<td>5065.41</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Analysis of Debilitating Anxiety Score by Group

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>5</td>
<td>78.74</td>
<td>15.75</td>
<td>.403</td>
<td>.847</td>
</tr>
<tr>
<td>Within Groups</td>
<td>239</td>
<td>9347.93</td>
<td>39.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>244</td>
<td>9426.67</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### One Way Analyses of Variance of Performance Measures by Groups

#### Analysis of Score on First Exam by Group

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>5</td>
<td>301.72</td>
<td>60.34</td>
<td>1.70</td>
<td>.135</td>
</tr>
<tr>
<td>Within Groups</td>
<td>239</td>
<td>8485.84</td>
<td>35.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>244</td>
<td>8787.57</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Analysis of Score on Second Exam by Group

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>5</td>
<td>152.71</td>
<td>30.54</td>
<td>.797</td>
<td>.553</td>
</tr>
<tr>
<td>Within Groups</td>
<td>239</td>
<td>8396.73</td>
<td>38.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>244</td>
<td>8549.44</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.

Correlations Between Anxiety Measures and Performance Measures

For the Class as a Whole

<table>
<thead>
<tr>
<th>Anxiety Measure</th>
<th>First Exam</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=245</td>
<td>n=225</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined Anxiety</td>
<td>-.167</td>
<td>-.124</td>
<td>(.004)</td>
<td>(.032)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilitating Anxiety</td>
<td>.0969</td>
<td>.065</td>
<td>(.065)</td>
<td>(.166)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debilitating Anxiety</td>
<td>-.163</td>
<td>-.126</td>
<td>(.005)</td>
<td>(.030)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For Each Group Individually

<table>
<thead>
<tr>
<th></th>
<th>FIRST EXAM</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety Measure</td>
<td>Comment A.</td>
<td>Comment B.</td>
<td>Hints &amp; All</td>
<td>Music A.</td>
<td>Music B.</td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td>n=35</td>
<td>n=34</td>
<td>n=50</td>
<td>n=45</td>
<td>n=40</td>
<td>n=41</td>
</tr>
<tr>
<td>Combined Anxiety</td>
<td>-.010</td>
<td>-.226</td>
<td>-.348**</td>
<td>-.044</td>
<td>-.172</td>
<td>-.237</td>
</tr>
<tr>
<td>Facilitating Anx.</td>
<td>.014</td>
<td>.155</td>
<td>.047</td>
<td>.069</td>
<td>.119</td>
<td>.272*</td>
</tr>
<tr>
<td>Debilitating Anx.</td>
<td>-.021</td>
<td>-.201</td>
<td>-.456**</td>
<td>-.011</td>
<td>-.166</td>
<td>-.144</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SECOND EXAM</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety Measure</td>
<td>n=34</td>
<td>n=33</td>
<td>n=46</td>
<td>n=43</td>
<td>n=35</td>
<td>n=34</td>
</tr>
<tr>
<td>Combined Anxiety</td>
<td>.142</td>
<td>-.102</td>
<td>-.201</td>
<td>-.149</td>
<td>-.127</td>
<td>-.131</td>
</tr>
<tr>
<td>Facilitating Anx.</td>
<td>-.002</td>
<td>.135</td>
<td>.020</td>
<td>.109</td>
<td>-.041</td>
<td>.120</td>
</tr>
<tr>
<td>Debilitating Anx.</td>
<td>.170</td>
<td>-.042</td>
<td>-.274*</td>
<td>-.140</td>
<td>-.202</td>
<td>-.103</td>
</tr>
</tbody>
</table>

* - \( p < .05 \)

** - \( p < .01 \)
determined in more detail by examining the analyses of variance which probed the effects of test anxiety, experimental group membership, and their interactions, on performance. In order to perform these analyses each group of students was divided into three, those highest on each scale of test anxiety, those lowest on each scale, and those who scored intermediately on each scale. (Note that the division points for the three groups for each scale were based on the scores of the class as a whole on each anxiety scale.) The means of each of these section in each group and for each of the anxiety scales are reported in Table 6.

The results of the analyses of variance are reported in Table 5. The particular results to be examined first are those included under 'main effects' and labelled with the name of the test anxiety scale in question. Note that according to these analyses the relationship between debilitating anxiety and performance on the second exam is not statistically significant, although it does approach statistical significance. All other results are consistent with the analyses of the correlations reported on Table 4. These results all indicate that test anxiety, as it was measured in these scales, does have a debilitating effect on student performance in an exam situation. These results also indicate that facilitating anxiety, at least as it was measured here, does not correlate with successful performance.

The major hypothesis of this experiment was that the various experimental conditions would act to alleviate some of the debilitating effects of test anxiety on performance. If this hypothesis were true, one would expect an interaction, statistically, between measured level of test anxiety and group membership, in their effects on performance. This
### Table 5.

**Analyses of Variance: Performance Scores by Group and Anxiety Measures**

#### First Exam by Group and Combined Anxiety

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>295.82</td>
<td>5</td>
<td>59.16</td>
<td>1.71</td>
<td>.133</td>
</tr>
<tr>
<td>Combined Anxiety</td>
<td>387.70</td>
<td>2</td>
<td>193.85</td>
<td>5.59</td>
<td>.005</td>
</tr>
<tr>
<td><strong>Two-Way Interactions of Anxiety and Group</strong></td>
<td>228.99</td>
<td>10</td>
<td>22.90</td>
<td>.66</td>
<td>.999</td>
</tr>
<tr>
<td>Residual</td>
<td>7869.16</td>
<td>227</td>
<td>34.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### First Exam by Group and Facilitating Anxiety

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>258.38</td>
<td>5</td>
<td>51.68</td>
<td>1.48</td>
<td>.196</td>
</tr>
<tr>
<td>Facilitating Anx.</td>
<td>118.73</td>
<td>2</td>
<td>59.36</td>
<td>1.70</td>
<td>.183</td>
</tr>
<tr>
<td><strong>Two-Way Interaction of Anxiety and Group</strong></td>
<td>331.03</td>
<td>10</td>
<td>33.10</td>
<td>.95</td>
<td>.999</td>
</tr>
<tr>
<td>Residual</td>
<td>7219.10</td>
<td>227</td>
<td>34.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### First Exam by Group and Debilitating Anxiety

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>270.84</td>
<td>5</td>
<td>54.17</td>
<td>1.62</td>
<td>.155</td>
</tr>
<tr>
<td>Debilitating Anx.</td>
<td>417.59</td>
<td>2</td>
<td>208.80</td>
<td>6.25</td>
<td>.003</td>
</tr>
<tr>
<td><strong>Two-Way Interaction of Anxiety and Group</strong></td>
<td>333.13</td>
<td>10</td>
<td>33.313</td>
<td>.99</td>
<td>.999</td>
</tr>
<tr>
<td>Residual</td>
<td>6918.14</td>
<td>227</td>
<td>33.42</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5. (continued)

Second Exam by Group and Combined Anxiety

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>152.65</td>
<td>5</td>
<td>30.53</td>
<td>.80</td>
<td>.999</td>
</tr>
<tr>
<td>Combined Anxiety</td>
<td>291.79</td>
<td>2</td>
<td>145.90</td>
<td>3.82</td>
<td>.023</td>
</tr>
<tr>
<td>Two-Way Interaction of Anxiety and Group</td>
<td>194.94</td>
<td>10</td>
<td>19.49</td>
<td>.51</td>
<td>.999</td>
</tr>
<tr>
<td>Residual</td>
<td>7910.00</td>
<td>207</td>
<td>38.21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Second Exam by Group and Facilitating Anxiety

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>147.32</td>
<td>5</td>
<td>29.46</td>
<td>.773</td>
<td>.999</td>
</tr>
<tr>
<td>Facilitating Anx.</td>
<td>37.08</td>
<td>2</td>
<td>18.54</td>
<td>.486</td>
<td>.999</td>
</tr>
<tr>
<td>Two-Way Interaction of Anxiety and Group</td>
<td>468.13</td>
<td>10</td>
<td>46.81</td>
<td>1.23</td>
<td>.274</td>
</tr>
<tr>
<td>Residual</td>
<td>7891.53</td>
<td>207</td>
<td>38.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Second Exam by Group and Debilitating Anxiety

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>141.01</td>
<td>5</td>
<td>28.20</td>
<td>.728</td>
<td>.999</td>
</tr>
<tr>
<td>Debilitating Anx.</td>
<td>193.83</td>
<td>2</td>
<td>96.91</td>
<td>2.50</td>
<td>.082</td>
</tr>
<tr>
<td>Two-Way Interaction of Anxiety and Group</td>
<td>186.28</td>
<td>10</td>
<td>18.63</td>
<td>.48</td>
<td>.999</td>
</tr>
<tr>
<td>Residual</td>
<td>8016.62</td>
<td>207</td>
<td>38.73</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Breakdown of Performance Scores by Group and by High, Medium, and Low Anxiety

Scores on First Exam

Note: The number of subjects in each category in each group is shown in parentheses beside the appropriate mean. The standard deviation of the exam score for each group is shown in parentheses below the means.

<table>
<thead>
<tr>
<th>Anx. Measure</th>
<th>Comment A.</th>
<th>Comment B.</th>
<th>Hints &amp; All</th>
<th>Music A.</th>
<th>Music B.</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Anx.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>27.6(9)</td>
<td>28.0(10)</td>
<td>26.9(13)</td>
<td>27.8(12)</td>
<td>27.0(13)</td>
<td>26.7(15)</td>
</tr>
<tr>
<td>Medium</td>
<td>26.9(14)</td>
<td>26.4(11)</td>
<td>25.2(21)</td>
<td>28.2(19)</td>
<td>28.9(13)</td>
<td>24.5(16)</td>
</tr>
<tr>
<td>High</td>
<td>27.7(12)</td>
<td>24.2(13)</td>
<td>21.3(16)</td>
<td>27.0(11)</td>
<td>23.9(16)</td>
<td>23.3(16)</td>
</tr>
<tr>
<td>Facilitating Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>28.3(9)</td>
<td>25.8(12)</td>
<td>24.9(19)</td>
<td>26.1(9)</td>
<td>23.9(16)</td>
<td>22.6(12)</td>
</tr>
<tr>
<td>Medium</td>
<td>27.2(18)</td>
<td>24.1(11)</td>
<td>24.4(12)</td>
<td>27.2(21)</td>
<td>28.8(12)</td>
<td>24.2(15)</td>
</tr>
<tr>
<td>High</td>
<td>26.5(8)</td>
<td>28.3(11)</td>
<td>23.9(19)</td>
<td>29.8(12)</td>
<td>27.2(14)</td>
<td>26.5(20)</td>
</tr>
<tr>
<td>Debilitating Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>27.3(9)</td>
<td>26.9(11)</td>
<td>27.3(13)</td>
<td>26.1(13)</td>
<td>28.7(11)</td>
<td>24.4(16)</td>
</tr>
<tr>
<td>High</td>
<td>26.6(13)</td>
<td>23.2(15)</td>
<td>21.4(17)</td>
<td>27.7(12)</td>
<td>24.7(19)</td>
<td>23.5(15)</td>
</tr>
<tr>
<td>S. D.'s</td>
<td>(5.7)</td>
<td>(5.8)</td>
<td>(5.5)</td>
<td>(6.3)</td>
<td>(6.0)</td>
<td>(6.5)</td>
</tr>
</tbody>
</table>
Table 6. (continued)

Breakdown of Scores on Second Exam by Group and Anxiety Level

<table>
<thead>
<tr>
<th>Anx._Measure</th>
<th>Group</th>
<th>Comment A.</th>
<th>Comment B.</th>
<th>Hints &amp; All</th>
<th>Music A.</th>
<th>Music B.</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>24.4(9)</td>
<td>27.2(10)</td>
<td>26.8(12)</td>
<td>28.3(13)</td>
<td>25.3(12)</td>
<td>25.6(12)</td>
<td>26.8(10)</td>
</tr>
<tr>
<td>Medium</td>
<td>26.5(13)</td>
<td>27.1(11)</td>
<td>26.9(20)</td>
<td>28.3(18)</td>
<td>29.3(11)</td>
<td>28.5(11)</td>
<td>28.3(9)</td>
</tr>
<tr>
<td>High</td>
<td>26.2(12)</td>
<td>24.3(12)</td>
<td>22.1(14)</td>
<td>26.7(12)</td>
<td>25.3(12)</td>
<td>25.7(12)</td>
<td>25.7(15)</td>
</tr>
<tr>
<td>Facilitating Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>26.0(9)</td>
<td>26.8(11)</td>
<td>26.6(18)</td>
<td>25.4(10)</td>
<td>25.2(12)</td>
<td>24.9(12)</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>26.2(17)</td>
<td>22.3(11)</td>
<td>24.2(11)</td>
<td>28.3(20)</td>
<td>28.5(12)</td>
<td>28.2(12)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>24.9(8)</td>
<td>29.3(11)</td>
<td>24.9(17)</td>
<td>29.3(13)</td>
<td>26.4(11)</td>
<td>27.3(12)</td>
<td></td>
</tr>
<tr>
<td>Debilitating Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>24.7(9)</td>
<td>26.2(11)</td>
<td>27.8(12)</td>
<td>28.4(14)</td>
<td>28.8(9)</td>
<td>28.4(11)</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>26.3(12)</td>
<td>28.0(8)</td>
<td>25.9(19)</td>
<td>28.4(16)</td>
<td>26.8(12)</td>
<td>26.1(10)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>26.2(13)</td>
<td>25.0(14)</td>
<td>22.8(15)</td>
<td>26.6(13)</td>
<td>25.3(14)</td>
<td>25.8(13)</td>
<td></td>
</tr>
<tr>
<td>S.D.'s</td>
<td>(4.7)</td>
<td>(6.4)</td>
<td>(6.4)</td>
<td>(6.2)</td>
<td>(6.8)</td>
<td>(6.4)</td>
<td></td>
</tr>
</tbody>
</table>
interaction can be measured using the analysis of variance. The analyses, reported in Table 5, show that there were no statistically significant interactions between any of the measures of test anxiety and group membership for either of the examinations.

Effects on Various Item Types

It was considered possible that the effects of the experimental conditions might manifest themselves on only certain types of questions in the examinations. Consequently, scores on the first examination were broken down into four types: test items which probed essentially factual material, test items which probed comprehension or ability to apply course material, test items which proved easy for most students, and test items which proved difficult for most students.

Test items which probed factual material and test items which probed comprehension and application were separated subjectively by the experimenter. Thirty-five items were labelled 'factual', seven items were labelled 'higher order', and eight items were considered too difficult to label and were ignored. (The separation was also done by three other graduate students. All three were more conservative and labelled many more items as 'factual'. In order to have an adequately large sample any item labelled 'higher order' by the experimenter and by one of the other three was included in the analysis.)

The classification of items into 'easy' or 'difficult' was done by item analysis. A median split was performed so that twenty-five items could be included in each classification of difficulty. (The median difficulty level was .53, that is, 53% of the students answered the median item correctly.)
The analyses of variance of each of these types of items by groups and by combined anxiety score are presented in Table 7. As with the analyses on the total score there were no significant interactions between the groups and level of test anxiety for any of these item types.

Note that the only result in any of these analyses which does not conform to the pattern of the analyses of the total score on the first exam is the result which indicates that the variance of score on 'higher order' items (those which probed comprehension and ability to apply course material) was not influenced by test anxiety score.

**Examination of Self-Report Measures**

While the major effort in this experiment was to assess changes in performance which could be attributed to the experimental conditions, an attempt was also made to assess directly the effects of the conditions on students' feelings of anxiousness during the exam also to ascertain whether the students attributed any effect to the conditions. The means of student responses to these questions can be seen in Table 1.

One way analyses of variance were performed on these measures to see if student responses varied from group to group. These can be examined in Table 8. It was found that there were no differences between the groups in the anxiety levels experienced, but there were differences between the groups in student reaction to the experimental conditions.

Now it must be noted that student reaction to the conditions was quite limited in both of the comment groups and in the control group, as students in these groups often said that they felt neutral about the conditions (which was scored as a 3) simply because they weren't sure just what the
Table 7.

Analyses of Variance: Scores on First Exam for Item Types Analyzed by Group and Combined Anxiety Score

**Factual Items (35)**

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>93.93</td>
<td>5</td>
<td>18.79</td>
<td>1.15</td>
<td>.336</td>
</tr>
<tr>
<td>Combined Anxiety</td>
<td>96.63</td>
<td>2</td>
<td>48.32</td>
<td>2.96</td>
<td>.053</td>
</tr>
<tr>
<td>Two-Way Interaction of Anxiety and Group</td>
<td>102.84</td>
<td>10</td>
<td>10.28</td>
<td>0.63</td>
<td>.999</td>
</tr>
<tr>
<td>Residual</td>
<td>2893.18</td>
<td>177</td>
<td>16.35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Higher-Order Items (7)**

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>6.38</td>
<td>5</td>
<td>1.28</td>
<td>0.740</td>
<td>.999</td>
</tr>
<tr>
<td>Combined Anxiety</td>
<td>7.68</td>
<td>2</td>
<td>3.84</td>
<td>2.22</td>
<td>.109</td>
</tr>
<tr>
<td>Two-Way Interaction of Anxiety and Group</td>
<td>10.85</td>
<td>10</td>
<td>1.09</td>
<td>0.63</td>
<td>.999</td>
</tr>
<tr>
<td>Residual</td>
<td>305.56</td>
<td>177</td>
<td>1.73</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Easy Items (25)**

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>55.72</td>
<td>5</td>
<td>11.14</td>
<td>0.796</td>
<td>.999</td>
</tr>
<tr>
<td>Combined Anxiety</td>
<td>125.60</td>
<td>2</td>
<td>62.80</td>
<td>4.49</td>
<td>.012</td>
</tr>
<tr>
<td>Two-Way Interaction of Anxiety and Group</td>
<td>69.33</td>
<td>10</td>
<td>6.93</td>
<td>0.49</td>
<td>.999</td>
</tr>
<tr>
<td>Residual</td>
<td>2477.40</td>
<td>177</td>
<td>13.99</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7. (continued)

**Difficult Items** (25)

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>44.76</td>
<td>5</td>
<td>8.95</td>
<td>.96</td>
<td>.999</td>
</tr>
<tr>
<td>Combined Anxiety</td>
<td>96.99</td>
<td>2</td>
<td>48.50</td>
<td>5.18</td>
<td>.007</td>
</tr>
<tr>
<td>Two-Way Interaction of Group and Anxiety</td>
<td>59.67</td>
<td>10</td>
<td>5.97</td>
<td>.64</td>
<td>.999</td>
</tr>
<tr>
<td>Residual</td>
<td>1657.14</td>
<td>177</td>
<td>9.36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8.

One Way Analyses of Variance of Questionnaire Data by Groups

### Anxiety Level in Second Exam

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>5</td>
<td>2.57</td>
<td>.51</td>
<td>.66</td>
<td>.660</td>
</tr>
<tr>
<td>Within Groups</td>
<td>197</td>
<td>153.41</td>
<td>.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>202</td>
<td>155.98</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Effect of Conditions

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>5</td>
<td>8.74</td>
<td>1.75</td>
<td>2.89</td>
<td>.016</td>
</tr>
<tr>
<td>Within Groups</td>
<td>197</td>
<td>120.55</td>
<td>.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>202</td>
<td>129.29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Reaction to Condition

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>5</td>
<td>10.79</td>
<td>2.16</td>
<td>2.50</td>
<td>.032</td>
</tr>
<tr>
<td>Within Groups</td>
<td>197</td>
<td>170.16</td>
<td>.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>202</td>
<td>180.96</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Others' Reaction to Condition

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>5</td>
<td>1.36</td>
<td>.27</td>
<td>.66</td>
<td>.660</td>
</tr>
<tr>
<td>Within Groups</td>
<td>197</td>
<td>81.64</td>
<td>.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>202</td>
<td>83.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
conditions were. Consequently, examination of the means of the responses to these questions allows the conclusion that the difference highlighted in the analysis of variance was between group three, which was the group in the second exam which got all the conditions, music, comments, and hints, and group four, which was one of the two groups in the straight music condition. Group five, the other music condition only group, scored midway between these two groups and was significantly different than the control group in reaction to the conditions.
Results - Supplementary Experiments

First Supplementary Experiment

Five measures were recorded for each subject in this experiment. These were: 1) the student's score on the test anxiety scale with both facilitating and debilitating scales combined; 2) the student's score on the facilitating scale only; 3) the student's score on the debilitating scale only; 4) the student's score on the first examination in the course; and 5) the student's score on the second examination. The means of each of these measures for each of the groups are listed in Table 9.

Because students were allowed to volunteer to be in the experimental groups, there may have been some differences between those students who chose to actively participate in the experiment and those who chose not to. Indeed, t-tests between these two sub-samples of subjects in the experiment (see Table 10) showed that, to a very statistically significant degree ($p < .01$), students who volunteered to actively participate in the experiment scored higher on debilitating anxiety, lower on facilitating anxiety, and in the direction of greater test anxiety on the combined anxiety scale.

The effects of the experimental conditions on student performance can be most easily examined by examining the correlations between test anxiety scores and performance on the two examinations for each group. These correlations are listed in Table 11. Note, the difference between the correlation of combined anxiety and performance on the second exam for the second experimental group (the one which got commenting as a condition on the second exam) and the same correlation for the control group; this difference is significant statistically ($p < .05$). Note also that the difference between the correlations of debilitating anxiety and
performance for the second exam for these same two groups is also statistically significant ($p < .05$). These significant differences would normally indicate that the experimental condition had had an effect. However, two facts indicate that this was not the case. First, the commenting condition showed no facilitative effect on the first exam when it was used with a different group, whereas group two almost showed an effect with a different condition. Second, the fact that there was actually a reversal of test anxiety effects with group two on the second exam rather than just a lessening of the effects seems to indicate that there were probably some inaccuracies in the test anxiety scores in that group.
Table 9.

**Means of Subject Measures - by Group**

(with standard deviations in parentheses)

<table>
<thead>
<tr>
<th>Measures</th>
<th>Group*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I (n=19)</td>
</tr>
<tr>
<td>Combined Anxiety (hi is high anxiety)</td>
<td>56.21 (10.9)</td>
</tr>
<tr>
<td>Facilitating Anxiety (hi is facilitating)</td>
<td>18.56 (5.6)</td>
</tr>
<tr>
<td>Debilitating Anxiety (hi is debilitating)</td>
<td>24.79 (6.2)</td>
</tr>
<tr>
<td>Score on First Examination</td>
<td>42.00 (5.8)</td>
</tr>
<tr>
<td>Score on Second Examination</td>
<td>38.66 (7.2)</td>
</tr>
</tbody>
</table>

* - **Experimental Group Conditions**

Group I - Comment condition for first exam.

Hints condition for second exam.

Group II - Hints condition for first exam.

Comment condition for second exam.

Group III - Hints condition for first exam.

Disbanded for second exam; took exam with control group.
Table 10.

*t*-Tests Comparing Control Group Subjects versus Other Subjects on All Anxiety Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group</th>
<th>n</th>
<th>Mean (s.d.)</th>
<th>T</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Anxiety</td>
<td>Control</td>
<td>157</td>
<td>48.78 (11.1)</td>
<td>-3.64</td>
<td>195</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>40</td>
<td>55.70 (9.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilitating Anx.</td>
<td>Control</td>
<td>157</td>
<td>19.78 (4.9)</td>
<td>2.55</td>
<td>195</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>40</td>
<td>17.58 (4.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debilitating Anx.</td>
<td>Control</td>
<td>157</td>
<td>18.56 (7.8)</td>
<td>-3.55</td>
<td>195</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>40</td>
<td>23.28 (6.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 11.

**Correlations Between Anxiety Measures and Performance Scores**

<table>
<thead>
<tr>
<th>Anxiety Measure</th>
<th>Group</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
</tr>
<tr>
<td></td>
<td>n=19</td>
<td>n=14</td>
<td>n=7</td>
<td></td>
</tr>
<tr>
<td><strong>FIRST EXAM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined Anxiety</td>
<td>Comments</td>
<td>-.308</td>
<td>-.045</td>
<td>-.269</td>
</tr>
<tr>
<td></td>
<td>Hints</td>
<td></td>
<td></td>
<td>- .271**</td>
</tr>
<tr>
<td>Facilitating Anxiety</td>
<td>Comments</td>
<td>.382*</td>
<td>.196</td>
<td>-.235</td>
</tr>
<tr>
<td></td>
<td>Hints</td>
<td></td>
<td></td>
<td>.226**</td>
</tr>
<tr>
<td>Debilitating Anxiety</td>
<td>Comments</td>
<td>-.202</td>
<td>.058</td>
<td>-.533</td>
</tr>
<tr>
<td></td>
<td>Hints</td>
<td></td>
<td></td>
<td>-.244**</td>
</tr>
<tr>
<td><strong>SECOND EXAM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined Anxiety</td>
<td>Comments</td>
<td>-.302</td>
<td>.274</td>
<td>-.369</td>
</tr>
<tr>
<td></td>
<td>Hints</td>
<td></td>
<td></td>
<td>-.259**</td>
</tr>
<tr>
<td>Facilitating Anxiety</td>
<td>Comments</td>
<td>.356</td>
<td>-.069</td>
<td>-.191</td>
</tr>
<tr>
<td></td>
<td>Hints</td>
<td></td>
<td></td>
<td>.222**</td>
</tr>
<tr>
<td>Debilitating Anxiety</td>
<td>Comments</td>
<td>-.216</td>
<td>.277</td>
<td>-.648</td>
</tr>
<tr>
<td></td>
<td>Hints</td>
<td></td>
<td></td>
<td>-.230**</td>
</tr>
</tbody>
</table>

* - p < .05  
** - p < .01
Results - Supplementary Experiments

Second Supplementary Experiment

Four measures were recorded for each subject in this experiment. These were: 1) the student's score on the test anxiety scale with both facilitating and debilitating scales combined; 2) the student's score on the facilitating scale only; 3) the student's score on the debilitating scale only; and 4) the student's score on the first examination in the course. The means of each of these measures for each of the experimental and control groups are listed on Table 12.

Because students were allowed to volunteer for the experimental groups, one must check to see if there were differences in test anxiety between those who chose to actively participate in the experiment and those who chose not to. In fact, t-tests between these two subsamples of subjects in the experiment (see Table 13) showed that students who volunteered to be in the experimental groups scored significantly higher (p < .01) on debilitating anxiety than those who chose not to be in them. There were no differences between groups on either combined anxiety or facilitating anxiety.

The effects of the experimental conditions on student performance can be most easily discerned by examining the correlations between test anxiety scores and examination score for each group. These correlations are listed on Table 14. Note that although there are no statistically significant differences between any of the groups, the correlation of debilitating anxiety and exam score is noticeably smaller in the 'music before' condition than it is in the control group. The 'z' of this difference is 1.08, which corresponds to a .15 probability that the difference is one due to chance.
### Table 12.

**Means of Subject Scores - by Group**

(with standard deviations in parentheses)

<table>
<thead>
<tr>
<th>Measures</th>
<th>Group*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>n=19</td>
</tr>
<tr>
<td>Combined Anxiety (hi is high anxiety)</td>
<td>48.21 (7.0)</td>
</tr>
<tr>
<td>Facilitating Anxiety (hi is facilitating)</td>
<td>21.47 (3.3)</td>
</tr>
<tr>
<td>Debilitating Anxiety (hi is debilitating)</td>
<td>19.68 (5.7)</td>
</tr>
<tr>
<td>Score on First Exam</td>
<td>38.05 (5.9)</td>
</tr>
</tbody>
</table>

* - **Experimental Group Conditions**

**Group I** - Music before but not during exam.

**Group II** - Music throughout exam.

**Group III** - Coffee and cookies available.
Table 13.

**t-Tests Comparing Control Group Subjects with all Other Subjects on All Anxiety Measures**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group</th>
<th>n</th>
<th>Mean (s.d.)</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Anxiety</td>
<td>Control</td>
<td>150</td>
<td>49.54 (9.5)</td>
<td>-1.56</td>
<td>187</td>
<td>.121</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>39</td>
<td>52.18 (9.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilitating Anx.</td>
<td>Control</td>
<td>150</td>
<td>19.33 (4.8)</td>
<td>-.60</td>
<td>187</td>
<td>.546</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>39</td>
<td>19.85 (4.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debilitating Anx.</td>
<td>Control</td>
<td>150</td>
<td>18.87 (6.5)</td>
<td>-2.76</td>
<td>187</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>39</td>
<td>22.03 (5.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 14.

**Correlations Between Anxiety Measures and Examination Score**

<table>
<thead>
<tr>
<th>Anxiety Measure</th>
<th>I (n=19)</th>
<th>II (n=5)</th>
<th>III (n=15)</th>
<th>Control (n=150)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Anxiety</td>
<td>-.132</td>
<td>-.782</td>
<td>-.207</td>
<td>-.267**</td>
</tr>
<tr>
<td>Facilitating Anxiety</td>
<td>.268</td>
<td>.728</td>
<td>.121</td>
<td>.198**</td>
</tr>
<tr>
<td>Debilitating Anxiety</td>
<td>-.005</td>
<td>-.438</td>
<td>-.279</td>
<td>-.245**</td>
</tr>
</tbody>
</table>

*a - Experimental Group Conditions*

- **Group I** - Music before but not during exam.
- **Group II** - Music throughout exam.
- **Group III** - Coffee and cookies available.

** - $p < .01$
Discussion

Explanation of Results

The hypothesis of this experiment was that certain experimental conditions might act to alleviate the debilitating effects of test anxiety on student performance on normal mid-term examinations. The data reported in the previous section indicate that, at least in this particular instance, the experimental conditions did not alleviate the effects of test anxiety. There are three general explanations which would account for this lack of positive results. It is possible that there was a major flaw in the overall design of the experiment, it is possible that in the execution of the experiment flaws in the presentation of each of the conditions might have led to a mitigation of any or all of their effects, and, of course, it is possible that the conditions, even in the best of circumstances, do not produce the alleviating effects that were expected. Each of these possibilities will be examined in turn.

Although possible, it is unlikely that the results, or lack thereof, were due to problems in the overall experiment. Random assignment of subjects to experimental groups seems to have provided representative samples both in test anxiety and in ability. There is no question that the students perceived the exam to be a 'real' one and the distribution of scores on both exams was low enough that there were undoubtedly no mitigating effects due to any lack of difficulty. Still, the effects of test anxiety seen in this experiment, even in the control group, were generally lower than one would normally expect (and were lower, although not significantly so, than in either of the two supplementary experiments). It is possible
that this overall lowering of the effects of test anxiety was due to the fact that all students took their exams in relatively small groups, even though these groups, at between forty and fifty people in size, were not really small enough for one to have expected any such mitigating effects.

There was some confusion involved for many students in finding the correct examination rooms. While this was not great, it is always possible that it added to feelings of normal test anxiety. Of course, this confusion should have also affected students in the control group, although it may have been mitigated to some extent for them as the proctor in the control group was the course instructor and his presence would have reassured the students that they were indeed in the correct room.

Still, these possibilities seem insufficient to explain the lack of results recorded in this experiment. It seems safe to say that the overall design of the experiment was basically adequate to its task.

It is considerably more likely that there were problems in the presentations of the individual conditions and that these problems were the cause of the ineffectiveness of these same conditions.

That there were problems in the presentations of the individual conditions seems most obvious when one examines the 'commenting' condition. In this condition students were given an extra sheet of paper on which to put any comments about the exam that they felt like making. This mode of presentation was different than either Calvin (1957) or McKeachie (1955), both of whom in their experiments left room on the exam sheet for students to make their comments. However, it was similar to Bucky's (1972) presentation, although in his experiment he used a number of pieces of paper so that he could examine effects over a number of subsections of the test he used.
The major effect of this difference in presentation mode was that many of the students left their commenting to the end of the exam rather than making comments as they went through the exam. Previous research on commenting suggests that the effects of commenting are cumulative and it seems clear that leaving all commenting to the end of the exam would mitigate much of the effect this condition would normally have.

Although all students were aware that their 'comment sheets' were part of the experiment, many of the students did not realize that it was the commenting which was the experimental condition. These students sometimes thought that what was wanted on the comment sheets were comments about the experimental condition. These students, then, were probably fairly upset about the fact that they were supposed to comment about something which they couldn't identify. Also, these students didn't think of the comment sheets as suitable places to make comments about the exam and therefore they couldn't really be said to have even been in a real commenting condition.

Another problem with the presentation was that the comment sheets were clearly identified as part of the experiment and so it was not clear to many students that their comments would ever be read by the instructor of the course. This would tend to reduce both the volume of the comments and any alleviating effects they would normally have.

Given all these problems associated with the commenting condition and attributable to flaws in it's presentation, it doesn't seem quite so strange that no alleviating effects were observed. This is important as commenting is the one experimental condition in this experiment which had been replicated successfully in other experiments. If the lack of results in this experiment were not attributable to problems in the presentation one might be led to
conclude that there were real problems in the overall running of the experiment and that one could not trust the lack of positive results in either of the other experimental conditions.

It seems less likely that the lack of results in the 'hints' condition can be attributed to problems in the presentation in this particular experiment. All students received the hints and understood that using them was supposed to reduce their own test anxiousness. The presentation of the 'coping model' was done by the experimenter and was seen as convincing by the other proctor present. What is more likely is that the calming effect of a coping model cannot be transmitted to fifty students at a time and that without this calming effect the usefulness of the hints is negligible.

It also seems unlikely the the lack of effects of the music condition can be attributable to idiosyncratic problems in this particular presentation. Of course, there are differences between the condition as it was done in this experiment and the condition as it was done originally by Stanton in Australia. The music was different, although an attempt was made to minimize this difference. Stanton used "the quieter movements from several of Mozart's symphonies" while in this experiment Mozart flute quartets and the quieter movements of Rimsky-Korsakov's Sherezade were used. Still, all of this music comes easily under the same classification of quiet classical background music. Similarly, there seems to be no essential difference between Stanton's use of four speakers attached to a stereo tape player and the use here of two speakers attached to a monophonic machine.

Nor can it be said that there were differences in the 'reality' of the testing situation. Stanton reports using "mid-year or end-of-term tests (which) were held as part of their normal programme" as was done here.
So, to the extent that is possible, the presentation of the music condition was similar to the original experiment with this condition. There may have been other differences, such as Stanton's own feelings about the experiment, but it must be assumed that these aspects of the presentation could never be exactly the same in any realistic exam situation. One must presume that any problems or deficiencies in this experimental condition were not the result of an inadequate presentation of the condition.

In this section two possible explanations for the disappointing results that were seen in this experiment have been examined. It has been concluded that it was unlikely that there was a flaw in the overall design of the experiment that was significant enough to have accounted for such a lack of results. Likewise in two of the three experimental conditions, the coping model/hints condition and the music condition, it is unlikely that one could explain a lack of results as being due to flaws in the presentations of these conditions in this particular experiment. Therefore, one must consider the possibility, at least in these two conditions, that the experimental manipulations are not, in and of themselves, powerful enough to cause any lessening of the normal debilitative effects of test anxiety.
Implications of the Results for Test Anxiety Theory

As was explained in the introduction of this experiment, there are a large number of overlapping theories which attempt to explain anxiety. The character of these theories makes it difficult to predict a subject's performance in a specific, actual situation. Consequently, it is difficult to confirm or deny the validity of any of these theories based on the results of one particular experiment. Nonetheless, an attempt will be made here to comment on aspects of the various theories that seem to be more, or less, valid or important given the results of this particular experiment.

The lack of results in the commenting condition underlines the fluid or transient nature of test anxiety in a testing situation. Test anxiety changes from one moment to the next in an exam. Both stimulus response theory and general arousal theory would predict a gradual lessening of anxiety over time, but the arousal theory better incorporates the need, as witnessed here, for the subject to engage in some activity for this lessening to occur. If one imagines that arousal can be used up (by commenting) then the need for commenting during the exam makes sense.

Of course, along with his general arousal theory, Epstein makes the case that anxiety is produced by a lack of an appropriate purposeful activity to reduce the perceived threat. It seems likely that commenting would provide a person with such a purposeful activity and it would therefore make a difference whether the commenting was done during the test or afterwards.

It is possible, although less likely, to interpret commenting as a method available to the student to reduce a test's threat to his own self-image. By commenting it can be argued that the student is explaining his
failure to answer a test question in terms which reflect on his knowledge or interpretation of the specific question. By so doing, he avoids having his failure to answer a question reflect on the larger, and more anxiety provoking, question of his competence or self-worth. The only problem with this interpretation is that it doesn't demand that a student comment immediately in order to reduce his feelings of anxiety. It would seem likely that if this explanation were true then just knowing that commenting could be done at the end should be sufficient to lessen the student's feelings of inadequacy and consequent anxiety.

An interesting problem in any discussion of test anxiety is how to reconcile theories which focus on threats to self-image and the like with theories which focus on the stimulus qualities of the testing situation. The question always becomes: "just what is it about a test that provokes an anxious response?". Clearly playing music at a test changes the specific stimulus qualities of the test, yet it seemed to have no effect on test anxiety. This would seem to lend credence to the cognitive theorists, except that one could say that music at an exam is well within the generalization boundaries of the stimulus qualities of an exam situation, and this explanation makes a lot of sense. Still, it does raise the very interesting question of whether there are any physical changes that one could make in an exam situation which would change the stimulus properties sufficiently to cause a reduction in test anxiety.

Playing music should have a calming effect and the fact that it didn't pose problems with the arousal explanation of anxiety. One must explain either why music did not calm the arousal of the students, or one must find a way in which music added to arousal in addition to subtracting from it.
The second solution seems more likely. Student reaction to the music was very fearful because most thought that it would disturb them. Therefore, students were aroused (to worry) by the music at the same time as the music was supposed to be calming them. Any close to reasonable trade-off between these two effects would have caused the results seen in this experiment.

If this explanation of the effect of music is true, then two implications may be drawn. First, if the students were told that the music would be turned off before their exam began then only the calming effect of the music should be noticeable, albeit at a lower level. Stanton (1975) reports doing this in his second experiment and reports success. It was also done in the second supplementary experiment here and the results, though not statistically significant, were in the direction expected.

The second implication is that the arousing effect of worry about the music should only occur the first time it is used and not on any subsequent occasions unless the music is really distracting. Since the repetition of music in this experiment was no more successful than the first time, one could assume that this is the case.

There is one question that must be asked about reaction to the music and that is whether or not an individual reacts alone or whether his reaction is influenced by the reactions of others around him. One statistically significant result of this experiment was that groups 3 and 4 reacted very differently to their conditions and, for the second exam, both groups were exposed to the music condition. Since both groups were exposed to the same music being played on the same equipment, one must assume that there was some feeling shared among the students in each group.
which influenced their personal reaction to the music. What is interesting is that this group influence was, to a large degree, unnoticed consciously by the students. The question probing the 'reaction of other people' was put into the questionnaire because it was felt that students would speak more openly about other students' misgivings than they would about their own. Nonetheless, in both groups students did not ascribe their own feelings to other students with the same intensity they themselves felt. So, whether the students really were unwittingly influenced by those around them, or whether the statistically significant effect was merely a chance aberration must be left as an open question.

So far reaction to the music has been discussed as though it were the same for all students. However, it should be noted that a number of idiosyncratic responses were noticed by the experimenter and it is always possible that this was more the rule than the exception. The two responses which come most readily to mind were both negative responses. Two students complained that the music made them actually frightened as it reminded them of the music played in their dentists' offices. This sort of reaction is consistent with a stimulus-response view of the situation and could have been significant. If two students actually noticed it, then probably many others felt it. The second idiosyncratic response came from three students all of whom were flutists. These students said that although they enjoyed the music immensely they found it very distracting as they tended to listen carefully to the flute solos to the detriment of their exam performance.

There is one factor that may explain why neither the music condition or the commenting condition achieved the expected results. Previous
research indicated that students working under a time limit on an exam were more affected by test anxiety than students not under a time limit. Even though there were officially no time limits in any of the groups, students in both the commenting and the music groups may have felt some time pressure. With speedy students leaving the exam room early, others may have worried that they would be told eventually that they must be finished by a time earlier than they would have wished. Given such an irrational fear, students in the commenting conditions would be worried that they wouldn't have time at the end to complete their comment sheet, and students in the music group might be worried that the calming effect of the music was slowing them down more than it was slowing others in the same group. While these fears may have been irrational, they may, in fact, have affected student performance negatively.
Suggestions for Future Research

In making suggestions for future research one is torn between the desire to clarify some of the theory behind test anxiety and the need to identify conditions that would alleviate some of its debilitating effects in real-life situations.

There are a number of prime candidates for future research on the theory of test anxiety. One could attempt any number of studies which would have as their goals the identification of the important stimulus qualities which define test anxiety provoking situations. In addition to physical stimulus qualities, one could attempt to isolate out the characteristics of particular tests which make them 'ego-involving'. One could attempt to physiologically measure arousal (or drive) and see if these measures correlate with the performance effects of test anxiety. One could attempt to identify different types of questions that were differentially affected by different levels of test anxiety, and one could see if these types matched any of the predictions made by the various theories. Once could attempt to derive an accurate measure of anxiety in the testing situation. Or one could attempt to differentiate various aspects of test anxiety to see which operate in which circumstances.

This writer has a bias towards examining test anxiety as a dynamic system. In this sort of analysis one looks for identifiable limited aspects of the situation which have reliable and quantifiable effects. Therefore, it is suggested that future theoretical research into test anxiety focus on limited and circumscribed, yet demonstrable, relationships between particular conditions and particular situations and the effects they have on performance. It would seem that such an approach would be
more helpful to the person trying to understand the effects of test anxiety than any attempt to bring all of the conditions that affect anxiety into one neat theoretical package.

However, this writer has an even greater bias towards research of a practical nature. It would seem more profitable at the present time to try to isolate conditions that would alleviate the effects of test anxiety. Any research of this type should be done in two distinct and separate stages. First, the experimenter should attempt to isolate a technique that works. Doing this necessitates running very large groups of subjects through a variety of conditions. These conditions should be suggested both by the theories of test anxiety and by the intuitions of the experimenters. Using very large numbers of subjects to study each condition is important for two reasons. First, some conditions may yield but small effects and large groups are necessary to observe these effects statistically. Second, in the first stages of implementing any type of anxiety reduction condition, one must expect problems which will tend to mitigate the effects of any conditions on test anxiety. Large numbers of students will allow effects to be seen even if they are attenuated by the circumstances of an experiment.

Only after a technique or condition has been reliably isolated should any attempt be made to see if it has any more than transient effects. By waiting until one can demonstrate initial effects reliably, one avoids the peril of discarding a useful technique as transient merely because it was implemented incorrectly on the second occasion of its use.

As for techniques worthy of future exploration commenting seems to head the list. Also, playing soft music before rather than during an exam
may be worthy of attention. It seems clear that most tests should not have
time limits, and one may want to examine the effects of allowing students
at least the feeling that they may bring in certain books or other aids to
an exam. It is quite possible that searching through one's notes may be
very effective as the sort of 'purposeful activity' which would help to
diminish test anxiety, even though, in some cases, no answers of any
importance would be allowed among the papers brought into the exam.

One unusual way of reducing test anxiety is undoubtedly worth experimen-
tal attention, although it may be difficult to isolate and study. The
first commenting group in the main experiment scored consistently (if not
statistically significantly) better than the other groups. The explanation
of this effect, if it indeed was such, might lie with the proctor who was
assigned to that group for both exams. This particular graduate student
is very casual, easy-going, and light-hearted. It is quite possible that
his own personal demeanour changed the stimulus properties of the exam in
such a way as to reduce the effects of test anxiety on his students.

To summarize, there are a number of changes one could make in normal
examination conditions which might serve to mitigate detrimental effects
of test anxiety. These changes should be explored experimentally using
very large groups of students and using repeated administrations of the
conditions. In this way it will be possible to isolate examination
conditions which yield consistent beneficial results. Once effective
changes can be isolated, one can return to more theoretical research on
the nature of anxiety and its relationship to test anxiety and have at
one's disposal a number of useful experimental manipulations, ones that
yield consistent and significant effects.
References


Gorusch, R. L. The general factor in the TAQ. *Psychological Reports*, 1966, 19, 308.


Appendix 1.

EXPERIMENTAL QUESTIONNAIRE

1. Nervousness while taking an exam or test hinders me from doing well.
2. I work most effectively under pressure, as when the task is very important.
3. While I may (or may not) be nervous before taking an exam, once I start, I seem to forget to be nervous.
4. In a course where I have been doing poorly, my fear of a bad grade cuts down my efficiency.
5. Nervousness while taking a test helps me do better. Always 1-2-3-4-5 Never
6. When I am poorly prepared for an exam or test, I get upset and do less well than even my limited knowledge should allow.
7. The more important the exam, the less well I seem to do.
8. When I start a test, nothing ever distracts me.
9. During exams or tests, I block on questions to which I know the answers, even though I might remember them as soon as the exam is over.
10. I find that my mind goes blank at the beginning of an exam and it takes me a few minutes before I can function.
11. I am so tired from worrying about an exam that by the time I start it I find that I almost don't care how well I do.
12. In courses in which the total grade is based mainly on one exam, I seem to do better than other people. Always 1-2-3-4-5 Never
13. I look forward to exams.
14. Time pressure on an exam causes me to do worse than the rest of the group, all else being equal.
15. Although "cramming" under pre-exam tension is not effective for most people, I find that if the need arises, I can learn material immediately before a test, even under considerable pressure, and successfully retain it to use on the exam.
16. I find myself reading exam questions without understanding them, and I must go back over them so that they make sense.
17. I enjoy a hard test more than an easy one.
18. The more important the exam the better I seem to do.
19. When I don't do well on a difficult item at the beginning of an exam, it tends to upset me so that Always 1-2-3-4-5 Never I block on easy questions later on.

Student Number ____________
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Always 1-2-3-4-5 Never
Appendix 2.

To other exam, my anxiety level in this exam was:

\[
\begin{array}{ccccc}
1 & 2 & 3 & 4 & 5 \\
\text{much} & \text{higher} & \text{the same} & \text{lower} & \text{much lower}
\end{array}
\]

In the experimental conditions:

\[
\begin{array}{ccccc}
1 & 2 & 3 & 4 & 5 \\
\text{helped me and} & \text{had no effect} & \text{And much}
\end{array}
\]

In reaction to the experimental conditions was:

From that I could tell, the reactions of other people around me to the experimental conditions were:
Please feel free to make any comments about the exam on this sheet.
YE OLDE TRIED AND TRUE TEST ANXIETY REDUCTION HINTS
with money back guarantee!!

1. Remind yourself periodically to stop thinking about yourself and to concentrate on the task at hand.

2. Think about aspects of the test that might be especially interesting to you.

3. Don't allow yourself to get flustered by errors and difficult items, but keep reminding yourself to work on the exam at a steady pace.

4. Force yourself not to think about other people and how they might be performing on the test.

Good Luck!