The Effects of Immediate Forewarning of Test Difficulty on Test Performance

CHARLES J. WEBER
Eastern Illinois University

GEORGE Y. BIZER
Union College

ABSTRACT. Whereas prior research has demonstrated that warning students of an exam’s difficulty well in advance of the exam enhances performance, the current research demonstrated the effect of such forewarning immediately before examination administration. Moments before taking an examination in a laboratory environment, participants were either (a) informed that the examination would be difficult, (b) informed that the examination would be easy, or (c) not given any information about the exam’s ostensible difficulty. Participants of low trait anxiety performed better when told the exam would be difficult than when told that the exam would be easy. Conversely, participants of high trait anxiety performed worse when told the exam would be difficult than when told the exam would be easy. Moments before taking an examination in a laboratory environment, participants were either (a) informed that the examination would be difficult, (b) informed that the examination would be easy, or (c) not given any information about the exam’s ostensible difficulty.

Key words: anxiety, assessment, performance

CONSIDER A SITUATION IN WHICH STUDENTS ARE SEATED in a lecture hall, ready to take a calculus exam for which they have studied for weeks. The course instructor, while handing out the examinations, off-handedly mentions that the exam is very difficult, and in fact expects that very few students will do well on it. What effect will such a warning have on these students?

To date, research on the effects of perceived test difficulty on test performance have not examined situations like this, but rather situations in which the forewarning comes well before the exam. Specifically, researchers have shown that such advance warning of difficulty leads to better exam performance. For example, Sax and Reade (1964) indirectly controlled the perceived difficulty.

This article is based on Charles J. Weber’s senior honors thesis in the psychology department of Eastern Illinois University.

Address correspondence to George Y. Bizer, Department of Psychology, Union College, Schenectady, NY 12308; bizerg@union.edu (e-mail).
of a final examination by manipulating the difficulty of earlier examinations. Students who performed well on difficult earlier examinations performed better on the final examination than did students who performed well on easier prior examinations. Sax and Reade attributed this effect to the fact that students who anticipated a difficult final exam were more motivated to study than were students who anticipated an easy final exam.

More recently, Foos (1992) revealed similar results in an experimental setting. Foos told participants that an upcoming test would be either easy or difficult and would be in either essay or multiple-choice format. After fifteen minutes of study, all participants took the same examination. Foos’ analysis demonstrated that participants who expected a difficult test performed better than did participants who expected an easy test. Similarly, participants who expected an essay test performed better than did students who expected a multiple-choice test. Like Sax and Reade (1964), Foos argued that the effect was due to enhanced motivation to study in the difficult and essay conditions. This enhanced study led, in turn, to better performance on the subsequent test.

In short, the literature suggests that if a student perceives an upcoming examination to be difficult, he or she is likely to do better on the exam. Although the underlying mechanism has not been empirically tested, researchers have argued that this is due to a heightened motivation to study. Often, however, the perception of test difficulty comes too late to affect motivation and subsequent study. For example, in the minutes or moments leading up to an examination, students may learn from colleagues in earlier sections that the test they are about to take was difficult. Similarly, it may be possible that the first several items on an exam will be difficult, leading students to perceive that the rest of the examination will be similarly difficult. Or, as posed earlier, a professor might inform the students that the examination will be difficult. In such cases in which motivation to study is clearly not possible, what might the effect of perception of test difficulty be on test performance? The current research was designed to test this notion.

The Current Research

Whereas prior research indicates perception of difficulty well in advance of an examination impacts motivation to study, perception of difficulty invoked just before an examination obviously cannot. Instead, we hypothesize that the perception of test difficulty just before or as an exam begins will impact the amount of anxiety that participants feel (e.g., Hong, 1999; Kim & Rocklin, 1994). It is in turn the differential levels of anxiety that influence test performance.

The body of literature on test anxiety is long and abundant (for reviews, see Hembree, 1988; Spielberger & Vagg, 1995; Zeidner, 1998). One reason why test anxiety has garnered so much research is that the construct often predicts performance on examinations. Simply put, the research has traditionally
shown that as test anxiety increases, performance on an examination decreases (e.g., Cassady & Johnson, 2002; Holroyd, Westbrook, Wolf, & Badhorn, 1978; Horn & Dollinger, 1989; McDonald, 2001). As such, the dominant current conceptualization of test anxiety is a linear model associating high anxiety with poor test performance.

However, a half-century ago, researchers presented several theories in contrast to such a linear model. In one paper, Sarason, Mandler, and Craighill (1952) proposed a model in accordance with the Yerkes-Dodson (1908) law suggesting that there may instead be a curvilinear relation between anxiety and test performance. These researchers posited that moderate levels of anxiety (conceptualized as the sum of state and trait anxiety) would lead to better test performance than would higher levels of anxiety. Based on Sarason et al.’s research, Alpert and Haber (1960) further argued that, depending on the amount and type, anxiety may be either facilitative or debilitative. Sarason et al. and Albert and Haber, to varying degrees, provided evidence that the relationship between anxiety and performance may not be as simple as is commonly conceptualized. However, the curvilinear relation suggested by these researchers has not garnered much empirical investigation. Indeed, in a recent review, Anderson and Sauser (1995) lamented, “it is disappointing that the ripples from the Alpert-Haber (1960) splash have all but disappeared.” (p. 19).

We propose that the splash made by Sarason et al. (1952) and Alpert and Haber (1960) may help predict the effect of immediate forewarning on test performance. Many researchers suggest that perceiving an examination to be difficult serves to heighten anxiety (e.g., Head & Lindsey, 1983; Hong, 1999; Hong & Karstensson, 2002). If the amount of anxiety a student experiences during an exam can be conceptualized as the sum of the student’s trait anxiety and the additional state anxiety that results from perceiving an exam to be difficult, some interesting predictions can be made. Specifically, such warnings of difficulty should lead students of high trait anxiety to perform more poorly than otherwise due to the resultant higher-than-optimal anxiety levels. On the other hand, such warnings should lead students of generally low anxiety to perform better due to the resultant moderate levels of anxiety. Therefore, our study tested the hypothesis that a warning of examination difficulty immediately before test administration would enhance the performance of low-trait-anxiety students but diminish the performance of high-trait-anxiety students.

Method

Participants

Sixty-two Eastern Illinois University (EIU) psychology students took part in the study in partial fulfillment of a course requirement.
Procedure

After participants entered the room and gave informed consent, we asked them to complete the 20-item trait-anxiety questionnaire (measuring level of dispositional anxiety) from the State-Trait Anxiety Inventory (STAI; Spielberger, 1983). We then told participants that they were about to answer questions from a prior administration of the Graduate Record Examination (GRE). We also told them that the GRE is an important examination and, as such, they should take the task seriously. We then told participants how difficult they could expect the examination to be, after which they completed the 20-item state-anxiety inventory (measuring momentary levels of anxiety) from the STAI (Spielberger).

We then distributed exam booklets to the students. Booklets were comprised of 10 multiple-choice items from a previously administered GRE.\(^1\) We told participants they could answer the questions in any order, and that they could check their answers, but they would only be provided with 10 min to complete the test. After the 10 min expired, we instructed them to put their writing utensils down. Finally, we distributed a brief questionnaire of ancillary measures.\(^2\)

Manipulation

By session, we randomly assigned participants to one of three instruction sets: We told participants that the examination would be (a) particularly difficult, (b) particularly easy, or (c) they were told nothing about the difficulty of the exam. Participants assigned to the “difficult” condition ($N = 18$) were told:

You’ll be answering some actual GRE analytical questions today. We expect that you will find this test to be very difficult to take. That is, we find that most EIU students do poorly when taking this exam. You will likely score very poorly on this test.

Those assigned to the easy condition ($N = 22$) were instead told:

You’ll be answering some actual GRE analytical questions today. We expect that you will find this test to be very easy to take. That is, we find that most EIU students do well when taking this exam. You will likely score very well on this test.

Finally, those assigned to the neutral condition ($N = 22$) were not informed about the difficulty of the exam.

Results

We then submitted data to a condition (difficult, easy, neutral) $\times$ median-split trait anxiety (low, high) analysis of variance (ANOVA) with number of correctly answered items serving as the dependent variable. Although there were no main effects of condition or trait anxiety ($Fs < 1, ns$) on number of correctly answered items, there was a significant condition $\times$ trait anxiety interaction, $F(2, 57) =$
4.96, \( p = .01 \). As shown in Table 1, post-hoc Least-Significant Difference (LSD) comparisons revealed that, among participants of low trait anxiety, participants told that the exam would be difficult scored significantly higher than did participants who were told that the exam would be easy. Scores of participants in the no-instruction control condition did not score significantly differently from either the easy- or difficult-instruction group. Conversely, among participants of high trait anxiety, participants who were told that the exam would be difficult scored worse than did participants in the other conditions.

Surprisingly, there was no effect of the manipulation on state anxiety scores \( F(1, 63) < 1, \text{ ns} \), nor was there a relation between state-anxiety scores and number of problems correctly answered (\( \beta = .11, \text{ ns} \)), although the correlation between state and trait anxiety scores was positive and significant (\( \beta = .70, p < .001 \)). Similarly, there was no relation between state anxiety and test performance (\( r = .01, \text{ ns} \)).

**Discussion**

The examination plays a large part in determining overall grades in many college courses. Research has shown that a variety of variables can play a role in the grade a student receives on a particular examination. Factors such as note-taking behaviors (e.g., Williams & Eggert, 2002), the number of assignments due near the exam’s administration (Lore, 1978), and the administration of prior unannounced quizzes (Graham, 1999) have all been shown to play a role in how well a student will perform on a given examination.

The current research adds to our understanding of test performance by demonstrating that not only can warning of test difficulty affect test performance if provided in advance of a test (Foos, 1992; Sax & Reade, 1964;), but also such a warning can impact performance even if provided immediately prior to test administration. Unlike previous research, we found no main effect of perception of test difficulty on performance, but rather an effect moderated by trait anxiety: Low-anxiety students performed better when told that the test would be difficult than when told that the test would be easy, whereas high-anxiety students per-

**TABLE 1. Overall Scores as a Function of Trait Anxiety and Condition**

<table>
<thead>
<tr>
<th>Trait Anxiety</th>
<th>Difficult</th>
<th>No Instructions</th>
<th>Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>8.11(_a)</td>
<td>6.70(_{ab})</td>
<td>6.09(_b)</td>
</tr>
<tr>
<td>High</td>
<td>5.00(_a)</td>
<td>7.31(_b)</td>
<td>7.18(_b)</td>
</tr>
</tbody>
</table>

*Note. Within rows, means with different subscripts are reliably different by Fisher’s least significant difference test comparisons (\( p < .05 \) for participants of low trait anxiety, \( p < .06 \) for participants of high trait anxiety).*
formed worse when warned that the test would be difficult than they did when told the test would be easy.

**Implications**

Our research has some important implications. First, it suggests that forewarning of difficulty well in advance of an examination and forewarning of difficulty immediately before an examination have distinct effects on test performance. Whereas a perception of test difficulty well in advance of a test can provide the extra motivation for a student to study hard, thus enhancing performance (e.g., Foos, 1992; Sax & Reade, 1964;), this research suggests that such a perception immediately in advance of a test may instead have a more complex effect on performance. Our evidence is consistent with the possibility that depending on a person’s underlying level of trait anxiety, the additional anxiety brought on by a warning of difficulty may either facilitate or debilitate performance.

Second, it provides evidence, albeit indirectly, for a curvilinear effect of test anxiety on performance (cf. Alpert & Haber, 1960; Ball, 1995). The traditional “linear” model of anxiety would have predicted that both high-anxiety and low-anxiety participants would perform more poorly when anxiety is raised due to a forewarning of test difficulty. This is because, according to the linear model, more anxiety necessarily means poorer performance. This is not what we found. On the contrary, the current research showed that low-anxiety participants actually performed better when warned that the test would be difficult, an effect inconsistent with the linear view of anxiety. Our research, therefore, provides support for the curvilinear conceptualization consistent with Sarason et al. (1952) and Alpert and Haber.

**Limitations and Directions for Future Research**

One important limitation of the current research is that we provided no evidence for our proposed mechanism—that state anxiety felt during the examination served to enhance or diminish test performance. Thus, although much prior research shows that perceiving a test to be difficult enhances anxiety (e.g., Head & Lindsey, 1983; Hong, 1999; Hong & Karstensson, 2002), we were unable to demonstrate an effect of the manipulation on state anxiety, and as such were unable to test whether such state anxiety underlies the effect.

The failure to find evidence of this mechanism may be due to one of three general reasons. First, it may be that state anxiety does mediate the effect of immediate forewarning on performance, but our measure of state anxiety did not successfully gauge participants’ levels thereof during the examination. Future research may better gauge the hypothesized mechanism by directly measuring a physiological indicator of anxiety during the examination administration rather
than indirectly measuring anxiety through a self-report measure before examination administration. Directly measuring heart rate (e.g., Meyerhoff, Oleshansky, & Mougey, 1988), blood pressure (e.g., Vassend, Halvorsen, & Norman, 1987), or immunoglobulin levels (e.g., Vassend & Halvorsen, 1987) may prove useful.

Conversely, it may be the case that, contrary to our hypotheses, enhanced anxiety is not responsible for the effects found in the current research. It may instead be that another single process mediates the effect. One possible mediator may be the effort that participants put forth when taking the examination. It may be that when told that an examination will be difficult, participants of low-trait anxiety put forth extra amounts of effort in attempts to succeed in the face of a potentially difficult task. Conversely, it may be that participants of high-trait anxiety withdraw effort in the face of such a potentially challenging task. To shield themselves against a potential blow to their self-esteem, these participants may engage in self-handicapping behavior, withdrawing effort in attempts to attribute failure to external rather than internal reasons (e.g., Jones & Berglas, 1978).

Finally, it may be that there is no single process at all that underlies the effect: It may instead be the case that two different processes—one for participants of low-trait anxiety and another for participants of high-trait anxiety—occur simultaneously. For example, telling low-trait-anxiety participants that a test is difficult may cause them to perceive the warning as a challenge, leading them to focus more carefully on the items. Conversely, telling high-trait-anxiety participants that a test is difficult may cause them to feel diminished confidence, leading to reduced test performance. Future research may assess whether either of these alternative hypotheses may best account for the phenomenon.

Another prospective arena for future analysis deals with research environment. For example, we conducted our research in a laboratory, using an examination that did not have an effect on student grades. Although we explained the importance of the GRE to the students, it may be that their overall anxiety levels were unrealistically low in such an artificial environment. As such, in a situation in which actual grades were at stake, the moderating effect of trait anxiety might be wiped out by the real-world stress and anxiety that comes with taking real examinations. Future research may determine the generalizability of the current effects.

Almost any student will agree that examinations are one of the most stressful parts of college life. The current research harkens back to the Yerkes-Dodson law (1908), showing that low-anxiety students benefited from an immediate forewarning of test difficulty, while high-anxiety students suffered from such a warning. Indeed, warning students that the test they are about to take is difficult, even if well-intentioned, may help some students, but harm others in the process. Along with prior research, the current research suggests that such warnings should not come immediately before the exam, but early enough so to give students time to study harder, rather than simply experience heightened anxiety as the exams are distributed.
NOTES

1. A pilot study indicated that the ten questions used were of moderate difficulty (the average number of total correctly answered questions was 7.05 out of 10, \(SD = 1.93\)). In addition, the average time needed to complete the ten questions was 9.29 minutes (\(SD = 1.90\)).

2. One ancillary measure, perceived difficulty of the examination, differed as a function of instructional set—\(F(2, 60) = 5.3, p = .006\); \(M_{\text{difficult}} = 3.33, SD = .77\); \(M_{\text{control}} = 3.04, SD = .71\); \(M_{\text{easy}} = 2.59, SD = .67\)—suggesting that the manipulation did indeed influence participants’ perception of test difficulty. Other measures were the amount of perceived stress participants felt during the exam, perceived importance of the exam, how well participants thought the exam would predict graduate school acceptance, likelihood of applying for graduate school in the future, high school and college GPAs, and whether participants felt that the instructions impacted their test performance. None of these measures differed as a function of the manipulation, and the primary analysis remained consistent when controlling for each (\(Fs < 1.1, ns\)). They are therefore not discussed further.

REFERENCES


Manuscript received July 26, 2004
Revision accepted for publication August 25, 2005