

## **CORRECTNESS AND CONFIDENCE: EVALUATING USABLE KNOWLEDGE DIFFERENCES BETWEEN MULTIPLE-CHOICE TESTS AND SHORT-ANSWER EXAMS**

**Kent Byus and Gregory S. Black, College of Business, Texas A&M University-Corpus Christi, 6300 Ocean Dr., Corpus Christi, TX 78412; (361) 825-2686, (361) 825-3619**

### **ABSTRACT**

Higher education assurance of learning techniques should be designed to assess whether lessons learned produce usable knowledge. While knowledge is often defined as "justified" information, it is the limiting of knowledge assessment to "correctness of answers" that is problematic in the development of complete learning assessment (Hunt and Furustig 1989). The purpose of this study is to continue research into the usability of knowledge as it appears to be influenced by testing formats. This article considers differences which appear to exist based on examination format, specifically differences between multiple-choice and short-answer tests. Results indicate that both the raw scores and self expressed student confidence with responses is affected by these two formats.

### **BACKGROUND**

Darling and Hammond (2003) assert that objective, multiple-choice instruments are the predominant type of exam. A recent survey found that a majority of marketing and management professors (56.32%) use multiple-choice exams in their classes to assess student learning (Black and Wingfield 2006). The popularity in usage is attributed to several factors, such as ease of administration, uniformity of interpretation, and more research based issues like reliability and validity. While extremely popular, multiple-choice exams are unable to completely assess the student's usable knowledge. To adequately address the expanded dimensions of usable knowledge, it is beneficial to explore and analyze differences between multiple-choice examinations and the more subjectively based short-answer tests.

Mottet and Beebe (2006) found that multiple choice exams are not good instruments to assess student abilities at actually giving speeches. In this case, multiple choice exams provided evidence that the students had the knowledge of how to give presentations and speeches, but not whether they were actually able to use the knowledge in the delivery of the presentation. In addition, Wingfield and Black (2005) found that course designs primarily using multiple-choice exams to measure student performance were not as effective as other

course designs. Other recent studies have found similar results in comparing multiple choice exams and more subjective forms of exams (e.g., Arthur et al 2002; Bacon 2003; Chan and Kennedy 2002; Krieg and Uyar 2001; Kuechler and Simkin 2003; Lawrence and Singhania 2004; O'Neill 2001).

Byus (2004) illustrated that merely being correct on any exam provides only one dimension of the usability of student knowledge. For knowledge to be completely usable, it is important for the learner to be both correct and confident of their correctness (Hunt and Furustig 1989; Byus 2004). To date, the analysis of this correctness-confidence relationship has been studied within the context of multiple-choice objective exam forms. Herein, the authors explore the usability of knowledge within more subjective short-answer exams and then compare differences in student correctness and confidence between the two more pervasive assessment instrument types. Both multiple-choice and short-answer item tests are widely used and universally acknowledged testing methods. Regardless of format, it is important for the educator to understand the relationship between learning and knowledge usability. These two often opposing influences (practical and academic) suggest that educators explore approaches to assess knowledge, which capture some usability criteria.

This paper continues work (Byus 2004) to explore an infrequently used knowledge assessment procedure that provides a better understanding of students' usable knowledge as measured by both the correctness of answers selected and the confidence expressed regarding answers selected in an academic setting. The authors will introduce a process and a model for such an assessment and will report on findings and implications as they regard differences between objective multiple-choice exams and subjective short-answer exams.

Important to this and other continuing studies is an acknowledgment of a basic assumption of knowledge usability. That assumption is that a necessary, but not sufficient condition of knowing something is that the person "in-the-know" must have confidence about that which is known (Ayer 1958; Hunt and Furustig 1989). Further, confidence influences the level to which an individual can

effectively and efficiently use his or her knowledge. Quine (1987) argued that the ability to measure whether an individual's belief (response) can be considered knowledge focuses on understanding that there exists benefit both in being correct and in being confident with one's correctness.

### **KNOWLEDGE AND CONFIDENCE**

Sveiby (1997, p. 37) defines knowledge as "a capacity to act." It is precisely the intangibility and invisibility characteristics of "capacity" that makes understanding this aspect of usable knowledge even more important when considering marketing students becoming critical thinkers. This analytical capability supports the benefit of measuring usable knowledge, a combination of correctness and confidence, in assurance of learning outcomes. Further, when considering knowledge, it must be understood that at some level, knowledge is relative and there is some practical and academic value in a student being technically incorrect, especially if there is some systematic process that provides insight into the otherwise inaccurate answer: the usability of the knowledge. For the purpose of this study, usable knowledge is defined in accordance with Hunt and Furustig (1989, p. 238) as "... beliefs that are correct and confidently justified." This definition also fits well in the evaluation of marketing students' performance.

### **MEASURING USABLE KNOWLEDGE**

Armed with the general understanding of the conditions of usable knowledge, it is important to develop the concept of certainty. Quine (1987,) stated the following. "Knowledge connotes certainty (but) what shall we count as certain? Knowledge applies only to true beliefs, and only to pretty firm ones, but just how firm or certain they have to be is the question" (p. 48). Including certainty into the assessment of knowledge requires educators to consider the confidence of the responder in at least as great a degree as one would assess the correctness of the student to the tested item.

Using simple one-dimensional logic, if the instructor observes an incorrect answer, the interpretation is that the person does not know the answer or is "uninformed." Such an inference does not necessarily portray a complete or accurate assessment of the student's knowledge. In the overall assessment of knowledge, the question must also be asked and answered: "If that student is extremely sure and still selects the incorrect answer, is that individual uninformed, misinformed, guessing, or just plain wrong?"

As argued in Byus (2004), to more fully assess knowledge, one must also consider the degree of confidence that a student expresses about responses provided and then analyze the relationship between the correctness and the confidence. A study using similar methods found that high-performing students were more confident of their answers on a multiple-choice exam than were lower performing students (Koku and Qureshi 2004). This study further found that the confidence levels of all students increased with easier questions and decreased with more difficult questions.

The Byus (2004) model suggests that being confident or unconfident with either correct or incorrect responses produces a set of four distinct levels of usable knowledge. First, the student is correct and confident and therefore is considered to be well informed. Second, the student is correct, but unconfident, which suggests that the student is in possession of partial information and incomplete knowledge. Third, a student providing an incorrect response and who is unconfident with the response produces a situation that suggests that the student is completely uninformed and that the level of knowledge possessed is "unusable." Finally, the student providing an incorrect response while being confident of the response's correctness is a situation referred to as being misinformed. To assist in this analysis, the Hunt and Furustig (1989) model can be reduced to a matrix that provides a practical diagnostic tool (see Figure 1).

### **A COMPARATIVE STUDY**

Assessment of knowledge can be generally characterized as the process of evaluating the student's ability to critically analyze a problem and produce, from their knowledge, a response that is designed to solve or address the issues that surround the problem. This assessment process occurs within a competing framework of examining student understanding and the need to develop higher order thinking and communications. Objective multiple-choice examinations are reasonably proficient in assessing student understanding, yet are deficient in assessing higher order thought and communication skills. Subjective short-answer exams are reasonably proficient in assessment of both. Unfortunately, short-answer response exams are more open to subjective interpretation (Matzen and Hoyt 2004) and are more difficult to score and therefore, are less used by instructors seeking administrative efficiency in the grading and assessment process.

In addition to being more difficult to score, subjective short-answer exams limit the form and content of student responses (Popham 2002). This truncation of form and content may further aggravate the added limitation of student expressed confidence with the response. Kubiszyn and Borich (2003) argue that short-answer and essay type examinations are most likely used to more assess knowledge comprehension in an application setting, which also might limit a student's ability to be confident of the response, regardless of correctness.

Where multiple-choice items predictably restrict student responses, such instruments may also produce greater confidence as a direct result of such constraint. While students might produce more confidence, they might also be producing less high-order thinking, comprehension, and ability to apply knowledge gained or expressed. Multiple-choice exam items also produce the ability to guess, while short-answer responses effectively remove, or at least greatly reduce, such guessing.

Herein, the authors report the preliminary results of ongoing research. The exam outcomes of a single Principles of Marketing class are reported. Both exams were administered to undergraduate students of a medium-sized regional university in the Southwestern United States. These students are enrolled and were tested during the Fall Semester of 2004. While admittedly small, the sample provides surprisingly, significant insights into analyzing the differences in the relationship of correctness and confidence using two different examination formats. The first examination consisted of 25 multiple-choice items and the second test consisted of 25 short-answer items. Both exams covered three chapters of the same textbook, using reasonably identical in-class instructional methodology.

A basic assumption used with these two exam forms is that all students enrolled and tested in a Principles of Marketing course, regardless of examination format, would possess the ability to express a similar amount of specific academic marketing knowledge. This assumption is made with the acknowledgement that perhaps certain undergraduate students would be repeating the course, or that individual students may have been exposed to some on-the-job marketing training. There also exists a possibility that some students may have enrolled in other business courses that included a marketing module. Finally, this assumption is made with the understanding that while seemingly insignificant to the instructor, subtle differences in instructional methodology might have been displayed between each of the two examination sections.

Students were uniformly instructed to provide two responses to each of the 25 questions. The first response requested a specific, direct-answer to the marketing subject addressed in the test, and the second responses requested the student to indicate level of "sureness" (confidence) with the direct-answer selected on the marketing issue portion of the question. Examples of both multiple-choice and short-answer items can be seen in Table 1. Both questions were followed by an identical confidence measure, as illustrated in the examples. Unlike previous exploratory studies of the knowledge usability question (e.g., Byus 2004), no further instructions were provided to the students. However, students were asked to be as honest as possible when providing the self-assessment of their confidence. Still, it is reasonable to assume that there exists an element of guesswork and/or a lack of student compliance/cooperation, or even integrity, in all tests by an undetermined proportion of students. This stated, the authors suggest that students are by and large honorable and will, when asked and motivated, provide honest responses to questions posed.

Specific hypotheses examined in this article are limited to some rather general questions. More specific research questions are being considered in the ongoing study, of which this is the initial findings. The specific hypotheses assessed in this study include the following.

- H<sub>1</sub>: Students will score significantly lower on short-answer examinations than on multiple-choice examinations.
- H<sub>2</sub>: Students will self-assess at significantly lower levels of confidence with responses provided to short-answer examination items than they will to multiple-choice examination items.

## RESULTS

The demographic characteristics and distributional significances are illustrated in Table 2. As presented in Table 2, estimates of the student demographic make-up and distributions suggest that the class followed historic enrollment patterns as estimated from previous university statistics. Using Chi-square tests of distributional uniformity, none of the demographic characteristics suggest any statistically significant departures from the usual and expected. This historic pattern of uniformity allows the hypotheses studied to be considered within reasonably usual conditions.

The Cronbach's alpha coefficient of reliability for the five-point Likert confidence scale was examined and computed to be acceptable for both the multiple-choice exam ( $\alpha = .831$ ) and the short-answer exam ( $\alpha = .884$ ). The combined reliability statistic for the scale was also computed and it proved to be significant also ( $\alpha = .899$ ). A randomly assigned split-half test of reliability for the multiple-choice exam items produced an equally acceptable set of reliabilities, with .663 for 13 items and .752 for the remaining 12 items. The split-half statistics for the short-answer exam confidence scale were computed as .752 for 13 items and .821 for the remaining 12 items. The correlation between the two halves of the two exams was computed to be an acceptable .705 for the multiple-choice exam and .813 for the short-answer exam.

The data shows that students scored significantly lower on the short-answer exam when compared to the multiple-choice exam format. Specifically, the mean score for the multiple-choice examination was .733 and the mean score for the short-answer exam was .647. The mean score was computed based on the number of correct responses provided. A paired sample t-test used to determine the statistical significance of differences between the two computed scores provided statistical support for H1 ( $t = 5.470, p \leq .01$ ). (See Table 3 for results of hypothesis testing.)

The data also illustrates that for both exam formats, students appear to be fairly unconfident with their responses. Specifically, the mean confidence score for the multiple-choice examination was 2.78 and the mean confidence score for the short-answer exam was 2.53. The scale used was scored as 0 = "Not

sure at all," 1 = "Very unsure," 2 = "Somewhat sure," 3 = "Very sure," and 4 = "Extremely sure." This low level of student confidence also supported the authors' initial impression that students would, at this early stage in their business coursework, be relatively unconfident with their responses. Using a paired sample t-test to determine the statistical significance of differences between the confidence levels associated with each test format, H2 is also supported ( $t = 3.922, p \leq .01$ ).

## CONCLUSIONS

There are substantially more questions that must be examined in the context of the "correctness and confidence" concept. This discussion paper is a continuation of research into the usable knowledge issue. As illustrated, students are more successful (score) and more confident being assessed using multiple-choice examinations. Further, it appears that students' estimates of the confidence are provided in a generally reliable manner. This information, when coupled with other assessment data, can be of substantial benefit to instructors seeking to evaluate more than a student's correctness of answers. For today's marketing educator, the ability to diagnose both the extent of student's knowledge and the extent of their confidence provides substantial potential for improving future marketers with curriculum that promotes higher levels of correctness and greater levels of confidence.

**Tables, Figures, and References Available on Request**