Comparative Studies of True/false, Multiple Choice and Multiple-multiple Choice

Beth A. Pinckney, Glenda M. Borcher and Edgar T. Clemens

Abstract

Studies compared the students' success and source of error on exams presented as true-false, single answer multiple choice, and multiple-multiple choice exam formats. The data would suggest that, although statistical differences were not observed between the proportion of correct answers and exam format, significant differences were detected in the source of error for incorrect answers. Students answering questions as true-false were less likely to mark a true question false, and more likely to mark a false question true, than were students answering the same question in either of the multiple choice exam formats. Differences were noted for gender, academic major, and final grade.

Pinckney and Borcher are undergraduate and graduate teaching assistants, respectively, and Clemens is a professor within the Department of Animal Science, University of Nebraska-Lincoln, Lincoln, Nebraska 68583.

Introduction

Sensitivity to, and fairness in, testing student learning requires an understanding of the examination instrument and how students of varied personality and learning style perceive exams (Wilen, 1987). Some exam formats are not looked on with favor, while others are (Cirn, 1987; Skipper, 1988; Borcher et al., 1992). Students often perceive True-False style exams as containing a measure of "trickery", having key words or phrases which make a particular question true or false. Single answer multiple choice questions assure the student that one answer is correct and provide them the opportunity to attack questions through the deductive process of elimination (Wilen, 1987; Lundeberg & Fox, 1991; Weimer, 1992). Furthermore, single answer multiple choice questions are viewed with considerably more favor than are the multiple-multiple choice style questions (i.e. those having one or more correct choice).

In a series of earlier studies, Borcher et al. (1992) and Hallman et al. (1992) compared the use of the "crossword puzzle" exam format with those of the same questions presented in the traditional "fill in the blank" format. These studies suggested that the crossword puzzle might well be the better of the two exam formats in enhancing student success, reducing the degree of variability according to student learning style and personality type, and receiving favor as the format of preferred choice by students.

Purpose

The following studies were designed to compare the students' success and source of error on exams presented as true-false (TF), single correct answer multiple choice (MC) and multiple correct answer multiple choice (MMC) exam formats.

Procedures

Two studies were conducted to determine the effectiveness of utilizing the TF, MC and MMC exam formats in assessing and assisting student learning.

Study One

The first study involved 104 undergraduate students enrolled in an introductory animal science course. Two exams were given during the semester in which students received 30 exam questions, one third (10 questions) of which were

ATI students learn, think, and write about the concepts and practical applications associated with personality typing and the MBTI has helped them appreciate the differences in people and increased their effectiveness in communication, decision making, and problem solving.

References

Barrett, L., Sorensen, R., & Hartung, T. (1985, March). Personality type factors of faculty and student implications for agricultural college teaching. *NACTA Journal*, 29, 50-54.

Barrett, L., Sorensen, R., & Hartung, T. (1987, December). Personality types of agricultural college students: Implications for teaching, retention and recruitment. *NACTA Journal*, 31, 14-19.

Sorensen, R. and Hartung, T. (1987, September). Student diversity and personality type. *NACTA Journal*, 31, 26-29.

Johnson, C. A. (1991). *Teacher know thyself.* Unpublished master's thesis, Ashland University. Ashland, OH.

Lawrence, G. (1982). *People types & tiger stripes: A practical guide to learning styles* (2nd ed.). Gainesville, FL: Center for Applications of Psychological Type.

Myers, I. B., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator.* Palo Alto, CA: Consulting Psychologists Press.

Myers, I. B., & Myers, P. B. (1980). *Gifts differing.* Palo Alto, CA: Consulting Psychologists Press.

Provost, J. A., & Anchors, S. (Eds.). (1987). *Applications of the Myers-Briggs Type Indicator in higher education.* Palo Alto, CA: Consulting Psychologists Press.

McCann, J. S., Heird, J. C., & Roberts, D. Y. (1989, December). Effective teaching methods for personality types of competitive judging team students and classmates in livestock and horse evaluation classes. *NACTA Journal*, 33, 5-8.

McCann, J. S., Heird, J. C., & Roberts, D. Y. (1991, March). Personality typing of students competing on nationally competitive livestock and horse judging teams. *NACTA Journal*, 35, 30-32.

Zimmerman, A. (1991, March). A capstone problem solving/systems course at a two-year technical college. *NACTA Journal*, 35, 26-29.

presented in each of the exam formats (i.e. TF, MC & MMC). In addition, the exams were randomly distributed such that one-third of the students (approx. 35) received exams in which the first 10 questions were in either the TF, MC or MMC format. Subsequent groups of 10 questions (i.e. questions 11 - 20, & 21 - 30) were also randomized such that one-third of the students received 10 exam questions in each of the formats. This process ensured that each student received 10 questions in each of the exam formats, and each group of 10 questions was presented to one-third of the class in each exam format.

The exams were structured such that each question on the TF exam appeared as the same question on the MC and MMC exam formats, as shown in Table 1. This question was identified as the "primary question" (PQ), and comparisons were made as to the students' ability to correctly answer this question, or portions of the question, relative to each format. The remaining portion of the MC and MMC questions (i.e. those not identified as the primary question) were designated as "secondary questions" (SQ) and comparisons were made, between the MC and MMC exam formats, in the students' ability to correctly answers the secondary questions.

Each exam was scored as to the percent of primary questions answered correctly. Furthermore, those primary questions answered incorrectly were identified as being either: a true question marked false (TMF) or a false question marked true (FMT), and recorded as the percent of the total true or false questions, respectively. The percent of correctly answered SQ (MC & MMC formats only) were also determined, as was the percent SQ incorrectly answered as TMF or FMT.

Students were identified according to class (freshman, sophomore, junior or senior), major (animal science, preveterinary medicine or non-animal science) and gender (female or male).

Study Two

The second study involved 64 undergraduate students enrolled in an upper level animal science - Anatomy/Physiology course. Thirty five of these students were enrolled in the course during the fall and 29 in the spring semester. The study was conducted within the laboratory section of the course throughout each semester with each student receiving seven laboratory exams. For each weekly exam, approximately one-third of the students received the questions in each format (i.e. TF, MC or MMC). Weekly exams were randomly distributed with the intent that each student would receive approximately one-third of the exams in each format.

Each exam was identified according to questioning format (TF, MC or MMC), chronological number (exam #1 - 7), and semester (fall or spring). All exams were scored according to the methods presented in study one. Students were identified according to class rank, major and gender as in study one. In addition the final grade of each student was also noted for comparative studies.

Statistical analysis

Data were analyzed for analysis of variance using a completely randomized design. Statistically significant differences were set at $P<0.05$. All statistical analyses were performed with the use of SAS (1986).

Table 1. Examples of the True-False, Multiple Choice and Multiple-Multiple Choice exam formats showing the "primary" and "secondary" questions.

Example, True - False Question;
True or False.
Stimulation of the right vagal nerve causes a dramatic slowing of the heart. (true)
Example, Multiple Choice Question;
Stimulation of the right vagal nerve causes a:
a. increase in cardiac out-put.
b. dramatic slowing of the heart. (true)
c. increase in atria-ventricle contractions.
d. decrease in the transition via the bundle of His.
Example, Multiple-Multiple Choice Question:
Stimulation of the right vagal nerve causes a:
a. decrease in the cardiac out-put. (true)
b. dramatic slowing of the heart. (true)
c. increase in the atria-ventricle contractions.
d. increase in the distance between the resting membrane potential and stimulatory threshold. (true)

*Answers underlined indicate the primary questions within each exam format. Correct answers indicated as being "true".

Results

Table 2 presents the mean values for correct and incorrect answers for each study and exam format. Statistical differences between exam formats were not detected ($P>0.05$) for the PQ correct answers in either study. However, statistical differences ($P<0.05$) were noted for the relative proportions of answers incorrectly marked as TMF and FMT (Table 2). Students in both studies were less likely to mark a false TF question "true", and more likely to mark a true TF question "false", than were students answering

Table 2. Mean (%) correct and incorrect answers for students taking true-false, multiple choice and multiple-multiple choice exam formats.*

	True-False	Multiple Choice	Multiple Multiple-Choice	SE**
Study One (n = 104)				
Primary Question				
Correct Answers	58.5	58.8	61.7	2.1
Incorrect Answers				
True Marked False	37.7[a]	56.1[b]	47.0[c]	2.9
False Marked True	49.3[a]	12.3[b]	26.6[c]	3.1
Secondary Questions				
Correct Answers		83.0[a]	73.9[b]	1.0
Incorrect Answers				
True Marked False		23.4	26.8	2.9
False Marked True		14.8[a]	25.1[b]	1.2
Study Two (n = 64)				
Primary Questions				
Correct Answers	77.0	76.2	74.9	1.0
Incorrect Answers				
True Marked False	18.9[a]	31.4[b]	28.7	1.6
False Marked True	28.1[a]	11.8[b]	17.6[c]	1.7
Secondary Questions				
Correct Answers		85.6[a]	71.1[b]	0.9
Incorrect Answers				
True Marked False		32.4	37.1	2.1
False Marked True		10.6[a]	20.4[b]	1.5

*Values within a row with different superscripts as statistically different ($P<0.05$).
**SE indicates pooled standard error.

the same question as MC or MMC. In addition, students answering MC primary questions were significantly less likely (P<0.05) to mark a false question true than were students answering the same question as a MMC.

When given the opportunity to choose the correct answers, as in the MC and MMC secondary questions, students had significantly more correct answers when taking the exam as MC compared to MMC (Table 2). Whereas statistical differences were not detected (P>0.05) for the TMF incorrect answers, students taking the exam as MMC were significantly more likely to have answers FMT.

Statistical differences (P<0.05) were detected between students' academic program and their success rate on TF exams, but not for the MC exam formats (Table 3). Scores of students majoring in Animal Science, and taking the exam as TF, were lower (P<0.05) than those for Pre-veterinary students and non-animal science majors in the introductory level course (study one), but were not different from those of either of these groups in the advanced course (study two). Statistical differences were not detected between majors for the MC and MMC exam formats.

Table 4 presents the results of the comparison of TF, MC and MMC exam formats according to the students' class rank. Statistical differences in scores based on class ranking were generally not detected for the varied exam formats. The one exception occurred in the introductory course (study one) for which a statistical difference (P<0.05) was noted between junior and senior students taking the MMC exam.

Gender differences were detected only in study one, and only for students taking the exam as MMC (Table 5). In this case, male students were significantly (P<0.05) more successful at deriving the PQ correct answer than were their female counterparts. For these questions male students had fewer questions that were TMF (44.7 vs 52.4%, male, female respectively, not shown in the table).

Table 6 presents the data arranged according to exam format (TF, MC, & MMC) and the students final grade within the course. Note these exams represented only 25% of the students' final grade. Students receiving a final grade of "A" in the course had significantly (P<0.05) higher TF and MC exam scores than all other final grade groups. "A" students also had a significantly higher MMC exam score than did the "C" and less than "C" students, but not those students receiving a final grade of "B". True-false exams provided the greatest range of exam scores for the final grade groups, while the MMC exam format provided the least (Table 6). Generally, those students receiving a final grade of "B", "C" or less than "C" did not differ in their success rate for MC or MMC exams when answering either the PQ or the SQ.

Table 3. Mean (%) correct answers for true-false, multiple choice and multiple-multiple choice exam formats according to the students' academic major*.

Study One	True-False	Multiple Choice	Multiple Multiple-Choice	SE**
Primary Question				
Animal Science (n=55)	49.4[a]	61.4	60.0	3.8
Pre-Vet Med (n=38)	59.6[b]	56.4	62.1	2.7
Non-Animal Science (n=11)	76.4[c]	64.2	63.6	6.1
Secondary Question				
Animal Science		82.4	73.3	1.9
Pre-Veterinary Med		83.0	73.8	1.4
Non-Animal Science		84.2	76.4	3.1
Study Two				
Primary Question				
Animal Science (n = 33)	78.4[ab]	75.9	75.2[ab]	1.5
Pre-Vet Med (n = 9)	82.5[b]	76.8	79.0[a]	2.6
Non-Animal Ag (n = 15)	73.7[ab]	78.6	74.9[ab]	2.1
Non-Ag (n = 7)	72.1[b]	74.1	69.6[a]	3.6
Secondary Question				
Animal Science		86.0	70.3	1.4
Pre-Vet Med		87.8	73.4	2.4
Non-Animal Ag		86.8	72.6	2.0
Non-Ag		80.4	67.9	3.1

*Values within a column with different superscripts are statistically different (P<0.05).
**SE indicates pooled standard error.

Table 4. Mean (%) correct answers for true-false, multiple choice and multiple-multiple choice exam formats according to the students' class rank*.

Study One	True-False	Multiple Choice	Multiple Multiple-Choice	SE**
Primary Question				
Freshman (n = 74)	55.3	56.3	61.2[ab]	2.6
Sophomore (n = 19)	65.2	63.5	61.0[ab]	4.7
Junior (n = 5)	64.0	68.0	52.0[a]	9.2
Senior (n = 6)	66.7	63.3	76.6[b]	8.4
Secondary Question				
Freshman		81.9	72.8	1.3
Sophomore		85.6	76.1	2.4
Junior		89.2	78.6	4.6
Senior		81.2	75.5	4.2
Study Two				
Primary Question				
Freshman (n = 2)	71.8	79.2	74.3	5.5
Sophomore (n = 19)	78.8	78.6	76.0	1.8
Junior (n = 34)	76.6	75.8	76.3	1.4
Senior (n = 9)	74.8	73.7	65.7	2.9
Secondary Question				
Freshman		79.5	68.1	4.9
Sophomore		86.5	71.4	1.6
Junior		85.6	71.5	1.4
Senior		87.6	68.1	3.2

*Values within a column with unlike superscripts are statistically different (P<0.05).
**SE indicates pooled standard error.

Discussion

A number of studies have been conducted on the relative value, and disadvantages, of True-False (TF) and Multiple Choice (MC) exam formats (Dolisky & Reid, 1984; Grosse & Wright, 1985; Kolstad et al., 1985; Toppino & Brochin, 1989; Thiede, 1991). While both formats contain an error component due to guessing (Grosse & Wright, 1985), the MC format appears to provide the student greater opportunity for answer "recognition" (i.e. knowing that one of the choices is correct and being able to see that answer) (Thiede, 1991). Grosse and Wright (1985) analyses suggest that TF scores carry the liability that "more true-false items are required to achieve the same reliability expected from 5-choice items" (i.e. single correct answer MC format). It is of interest to note that the MMC exam is, in fact, a series of TF questions placed in the MC format. The present study

provides a unique opportunity for direct comparisons of the exact same question (PQ) presented in TF, MC and MMC formats.

These data would suggest that, although statistical differences were not observed between the PQ correct answer and exam format, significant differences were detected in the source of error. Students answering questions as TF were less likely to have a TMF, and more likely to mark a FMT, than they were when answering those same questions as a MC or MMC. Ebel (1972) suggests that, "In the absence of firm knowledge a student seems more likely to accept than to question a declarative statement whose truth or falsity he must judge." This appears to hold true for the TF, but not for the MC or MMC exam formats, when consideration is given to the PQ. Errors associated with the MC and MMC formats indicated a greater tendency for more TMF than FMT. It is important to remember, however, that for the MC format, once a seemingly correct answer is selected all remaining choices are presumed false, thus greatly increasing the incidence of TMF. However, this is not the case for the MMC format in which each choice can be true or false. Ebel (1972) suggestion that students seek truth in answers becomes again evident in the increased incidence of FMT on the MMC - SQ comparisons.

Cirn (1986) and Thiede (1991) have shown that the MC format correlates more highly with the free-response test (i.e. short answer) than does the TF. Their studies demonstrate a relationship between answer recognition and the recall thought process. However, Lundeberg and Fox (1991) suggest that MC tests are not clear-cut measures of recognition, but require retrieval and sometimes interpretation and application of the information. It would appear from the present study that the MMC format, while providing for recognition, recall and interpretation of the test material, also informs students that more than one correct answer can be derived for a single question, an appropriate scenario in dealing with life's questions.

Gender differences, although suggested in other studies (Skipper, 1988; Anderson, 1989), were not generally apparent relative to exam format. Nor were differences readily apparent for comparisons based on students' major or class rank and the exam format. Probably the question of greatest interest to the student is "how does the test format affect their final grade in the course"? Statistical differences were not detected between exam format and final grade. However, the MC exam format provided the greater range of mean scores by final grade and the MMC the least. "A" and "B" students taking the exam as MMC tended to have lower mean scores, relative to those same students taking the TF or MMC format, "C" students maintained comparable scores in all three formats.

Table 5. Mean (%) correct answers for true-false, multiple choice, and multiple-multiple choice exam formats according to the students' gender*.

Study One	True-False	Multiple Choice	Multiple Multiple-Choice	SE**
Primary Question				
Female Students (n = 39)	59.3	64.3	54.5[a]	3.7
Male Students (n = 65)	58.1	56.4	64.9[b]	2.5
Secondary Question				
Female Students		84.4	71.6	1.9
Male Students		82.4	74.9	1.2
Study Two				
Primary Question				
Female Students (n = 25)	76.1	76.6	75.4	1.7
Male Students (n = 39)	77.9	75.8	74.5	1.3
Secondary Question				
Female Students		85.0	70.8	1.5
Male Students		86.2	71.4	1.2

*Values within a column with unlike superscripts are statistically different (P<0.05).
**SE indicates pooled standard error.

Acknowledgements

Appreciation in expressed to Ms. Linda A. Pavlish for the statistical analysis. Published as journal article No. 92-5, Nebraska College of Agricultural Sciences and Natural Resources.

References

Anderson, J. (1989). Sex-related differences on objective tests among undergraduates. *Educational studies in Mathematics*, 20:165-177.

Anderson, P. S. (1988). An educology of tasting : American student attitudes about test formats. *International Journal of Educology*, 2:143-184.

Borcher, G. M., Hallman, J. E. & Clemens, E. T. (1992). The crossword puzzle as a teaching/examination tool. *Agricultural Education Magazine*.

Cirn, J. T. (1986). True/false versus short answer questions. *College Teaching*, 34: 34-37.

Dolinsky, D. & Reid, V. E. (1984). Types of classroom tests: objective cognitive measures. *American Journal of Pharmaceutical Education*, 48:285-290

Ebel, R. L. (1972). *Essential of Educational Measurement*. Englewood Cliffs, NJ. (cited by Grosse & Wright).

Grosse, M. E. & Wright, B. D. (1985). Validity and reliability of true-false tests. *Educational and Psychological Measurement*, 45:1-13.

Hallman, J. E., Borcher, G. M. & Clemens, E. T. (1992). Student characteristics and the crossword puzzle as a teaching/examination tool. *NACTA Journal*, 36:44-47.

Kolstad, R. K. et al., (1985). Format-dependent selection of choices on MC and MTF test items. *Journal of research and development in education*, 19:27-31.

Lundeberg, M. A. & Fox, P. W. (1991). Do laboratory findings on test expectancy generalize to classroom outcomes? *Review of Educational Research*, 61:94-106.

Powell, J. L. & Gillespie, C. (1990). Assessment : all tests are not created equally. *Proceedings, American Reading Forum*. pg 13. Sarasota, FL.

Skipper, C. E. (1988). Gender differences in preservice teachers' preference for college learning activities. *Proceedings, American Educational Research Association*. pg 8. New Orleans, LA. (continued.)

Table 6. Mean (%) correct answers for true-false, multiple choice, and multiple-multiple choice exam formats according to the students' final grade.*

Study One	True-False	Multiple Choice	Multiple Multiple-Choice	SE**
Primary Question				
Grade of "A" (n = 12)	84.4[a]	84.7[a]	80.4[a]	0.9
Grade of "B" (n = 29)	76.7[b]	75.6[b]	73.3[ab]	1.0
Grade of "C" (n = 19)	66.7[c]	71.9[b]	69.7[b]	1.3
Less than "C" (n = 4)	67.6[c]	60.0[c]	68.6[b]	2.2
Secondary Question				
Grade of "A"		92.4[a]	76.4[a]	2.4
Grade of "B"		84.5[b]	67.8[b]	3.1
Grade of "C"		84.4[b]	67.6[b]	2.9
Less than "C"		80.0[b]	67.6[b]	2.7

*Values within a column with unlike superscripts are statistically different (P<0.05).
**SE indicates the pooled standard error.