

Investigation of Probabilistic Multiple-Choice in a Structural Design Course

Major Adrian Owen Biggerstaff P.E., U.S. Military Academy

Major Adrian Biggerstaff is an Assistant Professor at the U.S. Military Academy, West Point, NY. He received his B.S. from the United States Military Academy, and M.S. degrees from Stanford University and Missouri University of Science and Technology. He is a member of ASEE and is a registered Professional Engineer in Missouri. His primary research interests include sustainable design, construction engineering, and engineering education.

Lt. Col. Brad Wambeke P.E., United States Military Academy

Lieutenant Colonel Brad Wambeke is the Civil Engineering Design Group Director at the U.S. Military Academy, West Point, NY. He received his B.S. from South Dakota State University; M.S. from the University of Minnesota; and Ph.D. from North Carolina State University. He is a member of ASEE and is a registered Professional Engineer in Missouri. His primary research interests include construction engineering, lean construction, and engineering education.

Investigation of Probabilistic Multiple-Choice Assessments in a Structural Design Course

ABSTRACT

Student assessment is a critical part of the learning process. Engineering courses often use objective student assessments to promote impartiality and grading efficiency, but many objective assessment methods do not provide educators sufficient information on their students' level of knowledge. The probabilistic multiple-choice question is a type of objective assessment that uses a 'reproducing scoring system' which encourages students to 'reproduce' their true knowledge state. Presently, the authors are unaware of any attempts to use probabilistic multiplechoice assessments in undergraduate civil engineering courses. This paper examines the impact of using probabilistic multiple-choice questions on students in CE404: Design of Steel and Wood The paper explores whether or not probabilistic multiple-choice assessments Structures. positively impact students' level of critical thought, their understanding of technical language used in the course, their ability to self-assess their level of understanding and confidence, and the propensity for self-learning. Student performance on several assignments is compared to the previous year when probabilistic multiple-choice was not used While there is no clear indication that probabilistic multiple choice had a significant positive impact in most of the areas, students using probabilistic multiple choice did score 4.8% higher overall on major graded events.

INTRODUCTION

Assessments inform students on their scope and depth of knowledge and serve as an indicator of teaching effectiveness for educators. Assessments can also help students think about course material in a different way ⁽¹⁾. However, assessments require time on the part of students and the educator. This paper examines the use of probabilistic multiple choice assessments in CE404: Design of Steel and Wood Structures, a required course for civil engineering majors at the United States Military Academy (USMA). CE404 builds on students' understanding of statics, mechanics, and structural analysis to design tension, compression, bending, and beam-column structural members. Students complete seven homework assignments, two mid-term exams, one Engineering Design Problem (EDP), and a final exam during the course. Until 2015, the student's homework assignments consisted primarily of free response design or analysis problems using a problem solving format.

Time is the most demanded resource for cadets at the USMA. A USMA Senior takes an average of 20 academic credit hours per semester not counting an additional military science course, a physical fitness course, mandatory participation in a competitive sport, and leadership responsibilities in the Corps of Cadets. In addition to the mandatory demands on a cadet's time, the typical cadet is also involved with club events, social media, or other extracurricular activities. The cadets' time demands create challenges in achieving the typical 1:2 ratio of

contact hours to out-of-class hours for an academic course. From 2005 to 2014, out-of-class time survey data indicates that cadets spend on average 60 minutes out of class for every hour in class, half the expected time. The time cadets do spend out of class on homework often involves repeating steps from in-class examples with different values to practice the "problem solving format." These homework problems are valuable to reinforce equations taught in class and to exercise the cadets' ability to communicate in writing, but these types of homework problems rarely go beyond the cognitive domain of *applying* ⁽²⁾.

The nature of structural member analysis and design is often repetitive and students are often able to follow a set of prescribed steps to arrive at the correct answer with minimal understanding of the structural analysis or design process. The concern is that while the student can follow the steps, more like a technician than an engineer, they are not internalizing the depth of knowledge required for understanding and solve problems of greater complexity that they may see in their careers. The regurgitation of in-class problem steps is due in part to the limited time cadets allocate for their out-of-class assignments.

Creating homework assignments that challenge students to think critically and do not exceed the expected out-of-class time is difficult. Additionally, free response questions that do challenge the students' comprehension is burdensome to grade when providing meaningful feedback. USMA does not employ Teaching Assistants, and instructors are required to manage course administration, write the course assessments, teach the material, provide office hours, and grade all course assignments. The time requirement is compounded when providing feedback to students with a poor understanding of the material.

ASSESSMENTS

Assessments serve a critical function in the education process ⁽¹⁾. A high-quality assessment confirms what the student knows, identifies the areas a student is weak, pulls together what the student has learned in class, exercises higher-order cognitive abilities, and provides a comparative assessment among students. A high-quality assessment also provides the instructor insight on their students' knowledge states ⁽³⁾.

Science, Technology, Engineering, and Mathematics (STEM) courses gravitate towards objective assessments to test student knowledge. Objective tests fit into two broad categories: selection-type (multiple-choice, matching, etc.) and supply-type (short answer, problem set, etc.) ⁽⁴⁾. Selection-type testing, specifically multiple-choice (MC) assessments, offer a variety of different methods for assessing student learning. These methods include, but are not limited to, the conventional MC test, elimination testing, confidence marking, probability testing, and two-stem questions. Many of the methods are structured to both discourage student guessing and provide opportunities to capture partial knowledge ⁽⁵⁾. MC testing facilitates objective, high-speed grading with the functionality to assess higher level thinking ⁽⁶⁾. Research has also found that modifying conventional MC questions can require students to more critically examine the

material being tested ⁽⁷⁾. The main limitations to MC assessments are the test writer's creativity, the tools available for grading, and the minimal exercise of student communication skills.

PROBABILISTIC MULTIPLE CHOICE

Probabilistic MC assessments use a 'reproducing scoring system' that provides a valuable service to students and educators. Reproducing scoring systems, also known as a "proper scoring systems," "admissible scoring systems," or "scoring systems which encourage honesty," create a system of rewards that incentivizes students to report their true belief or knowledge state ^{(8) (9)}. These systems provide the educator with greater insight into their students' true knowledge state and encourage students to spend more time assessing how well they understand the material ⁽¹⁰⁾.

The conventional single-answer MC question can mask a student's knowledge state from both themselves and the educator ⁽¹¹⁾. These assessments provide binary or dichotomous feedback: students get the right answer and full credit or the wrong answer and no credit. Students with low knowledge states who randomly guess on a four-answer-choice question have an expected score equal to 25% of the total points. If the students can eliminate one or two answer choice(s), their expected score moves towards 50% of the total points. The downside for guessing is earning zero point while the upside is getting full credit. The mutually exclusive and collectively exhaustive nature of conventional MC questions provide students little incentive to critically exam all answer choices once their 'correct' answer is identified.

A modification to this standard MC assessment is removing the mutual exclusivity of the problem by allowing more than one correct answer choice. A four-answer-choice problem with multiple correct answers decreases a student's chances of randomly selecting the correct answer from a probability of 0.25 to 0.067. The student has a much greater incentive for applying critical thought to each answer choice; however, while the multiple answer modification can capture partial knowledge and increase critical thinking, the assessment technique does not guarantee quality feedback on the student's knowledge because the educator does not know the student's degree of uncertainty for their selected answers ⁽⁹⁾.

Ease of computer access and advancements in computer applications are significantly improving educators' ability to assess student learning through MC assessments. Computer labs and personal computers provide students a venue for submitting out-of-class assignments, and course websites and assessment software provide instructors an opportunity to expeditiously collect and grade student work. These computer-based assessments can incorporate reproducing scoring rules that award credit for partial knowledge on MC questions ⁽³⁾ ⁽¹²⁾. This type of MC assessment requires students to assign a probability to each answer choice being correct. There are various graphical, qualitative, or numeric methods in existence for students to report their probabilities ⁽³⁾ ⁽¹²⁾ ⁽¹³⁾ ⁽¹⁰⁾.

Reproducing scoring systems can vary in both how they award points and the range of points they award. A spherical scoring rule awards points *locally* or based only on the probability

assigned to the correct answer; quadratic and logarithmic scoring rules can award points *globally* or based on the probability vector assigned across all answer choices. These different scoring rules also offer different bounds on the number of points a student can lose for assigning a low probability to the correct answer ^{(9) (10)}. For example, a spherical scoring rule will not award a student less than 0 points where as a logarithmic scoring rule can award a student up to $-\infty$ points for assigning zero probability to the correct answer.

Doctors, meteorologist, intelligence officers, and many other professionals must become comfortable reporting their level of certainty in the form of probabilities ^{(11) (14)}. Engineers also face great uncertainty in their profession, and there is great value in gaining practice reporting their level confidence as students. Research indicates that confidence reporting improves with practice and that higher levels of reported confidence indicate greater retention in knowledge or skill ^{(13) (15)}. Additionally, probability reporting vastly increases the amount of information the instructor gets on a student's knowledge state ⁽⁹⁾. This additional insight can aid the instructor in finding and correcting weaknesses in their students' understanding.

Probabilistic MC can assist in communication training by exercising the student's ability to concisely express subjective uncertainty, perform operational self-rating of confidence, and produce usable information for quantitative models ⁽⁹⁾. However, moving towards MC assessments and away from free response problems reduces the students' ability to convey a neat, methodical approach to solving a complex problem that requires the explanation of facts and assumptions used.

USE OF PROBABILISTIC MULTIPLE-CHOICE IN CE404

This study used numeric probability reporting to assess the students' knowledge state. Students assigned a total of 100% probability across the four answer choices and were awarded points locally (based only on the probability of the correct answer) using a logarithmic scoring rule shown in Equation 1 where A is the available points for the problem and p is the student's assigned probability to the correct answer. Logarithmic scoring rules are typically not bounded so students can in theory earn $-\infty$ as seen in Figure 1, which uses an A value of 5 points. Figure 1 also illustrates that a student that distributes their probabilities uniformly across all answer choices, a 0.25 probability in this case, earns zero points. A uniform distribution is an admission of ignorance, and the scoring rule awards no points. However, receiving zero points is a better outcome than guessing, which could result in an unrecoverable loss of points for students who are risk adverse or risk neutral.

Awarded Points =
$$A\left(1 + \frac{\ln p}{\ln 4}\right)$$
 (1)



Figure 1: Awarded points based on assigned probability to correct answer choice for a four-answer problem worth 5 points

This study bounded Equation 1 by normalizing the final grades to prevent any students from getting less than zero points on the probabilistic MC portion of the assignments. The grader first totaled the points each student earned in the probabilistic MC portion of the assignment. The student who earned the highest overall score on the questions received maximum points for that portion of the problem set. The student who received the lowest number of point, typically a large negative value, received zero points for that portion of the problem set. The remaining students earned a normalized score between 0 - 100% based on their performance relative to the highest and lowest scoring student.

The seven homework assignments or "problem sets" assigned to the 2016 CE404 students were very similar to the assignments given to 2015 CE404 students except for a portion of the problems were converted from free response to probabilistic MC assessments. Additionally, a few conceptual problems were added to the problem sets. Between 28 to 38 % of the available points in the problem sets were earned through probabilistic multiple choice, the remaining points were earned through the traditional free response questions that tested the problem solving format. Free response questions remained in the problem sets to exercise the students' ability to communicate technical work in writing.

The 2016 students were also assigned a "Problem Set 0" on the first day of class that reviewed statics in order to familiarize the class with the probabilistic MC assessment method.

Figure 2 provides an example probabilistic MC question and the points associated with different probabilities assigned to the correct answer. The "Points Awarded" values in Figure 2 have not been normalized in accordance with the previously described procedure.



Figure 2: Example Probabilistic Multiple Choice Question with Example Answer and Point Values

The same instructors taught the 2015 and 2016 CE404 students using the same board notes, course schedule, and labs. Additionally, all 2016 graded events were scored using the same scoring rubric used in 2015, the mid-term exams were identical to the 2015 exams except for minor changes in member dimensions and load magnitudes, and the 2016 end-of-term exam was identical to the 2015 exam and graded by the same instructors using the same scoring rubric.

DESIRED OUTCOMES

The goal of using probabilistic MC assessments in CE404 was to determine if the technique increased the student's:

- 1. level of critical thought throughout the course
- 2. understanding of technical language used in the course
- 3. ability to self-assess their level of understanding
- 4. ability to express their level of confidence
- 5. self-learning
- 6. quantitative performance on mid-term and end-term exams

This study focused on measuring the level to which the use of probabilistic MC affected the six objectives above through comparing mid-term and end-of-course exam scores between the 2015 and 2016 classes, giving the 2016 class the survey shown in Figure 3, comparing the end-of-course feedback between the 2015 and 2016 CE404 students, comparing the end-of-course

feedback across all civil engineering courses in 2016, and comparing the end-of- course feedback and course performance between the 2015 and 2016 CE404 student's in CE403: Structural Analysis. CE403 is the feeder class into CE404, and the CE403 data provides interesting comparisons because the two year groups' survey responses, out-of-class study time, and course performance between the two courses. The CE403 data is available in Enclosure 1.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
My level of critical thought increased as a result of the probabilistic multiple choice questions.	1	2	3	4	5
My understanding of the technical language in CE404 increased as a result of the probabilistic multiple choice questions.	1	2	3	4	5
My ability to self-assess my own knowledge increased as a result of answering the probabilistic multiple choice questions	1	2	3	4	5
I feel more able to express my level of confidence based on my current level of information (state of information) as a result of the probabilistic multiple choice questions.	1	2	3	4	5
The probabilistic multiple choice questions heightened my self-learning when my confidence was low on a question.	1	2	3	4	5

Figure 3: 2016 CE404 Student Survey taken during last week of class

RESULTS

The study provided both a formative and summative assessment of using probabilistic multiple choice. The formative assessment is primarily captured through the survey and questionnaire feedback provided by students and offers insight into the learning process. The summative assessment is captured through the performance on graded events and offers insight into the students' performance after the instructional process. The study results are organized by the study's six objectives, but it is important to note that the student survey data is not always consistent with other metrics. The authors contribute at least some of the discrepancies to the students' general sentiment towards the probabilistic assessments. Through formal and informal feedback, many students reported that they did not appreciate losing points due to selecting the incorrect answer. This inherent consequence in the probabilistic MC method used in CE404 may have tainted some of the students' views towards the assessment technique, which is reflected in the student survey data.

1. Increases the student's level of critical thought throughout the course

As seen in Figure 4, only 29% of the class felt as though probabilistic multiple choice questions increased their level of critical thought while 39% of the students felt the questions did not increase their level of critical thought. However, comparing the end-of-semester feedback between the 2015 and 2016 class in Figure 5 indicates a 3% increase in the students reported critical thought. The 2016 class had previously reported a 5% decrease in critical thought compared to the 2015 class in CE403's end-of-course feedback.

Additionally, comparing the end-of-semester feedback to all other courses in the civil engineering division, civil & mechanical engineering department, and the entire academy in Figure 6, critical thinking and reflection was clearly the highest in CE404.



Figure 4: CE404 2016 Student Survey Results for Question 1



Figure 5: CE404 2015 and 2016 end-of-semester feedback comparison



Figure 6: CE404 2016 end-of-semester feedback comparison to Civil Engineering (CE) Division, Civil & Mechanical Engineering (C&ME) Department, and United States Military Academy (USMA) courses

While the total out-of-class time was not a consideration in this research, Figure 7 does indicate that the 2016 CE404 students spent an average of 28 more minutes between lessons working on the CE404 material compared to the 2015 students. While reporting more time in CE404 than the 2015 students, the 2016 CE404 students reported spending 10 minutes less per lesson than the 2015 students in CE403. Since the scope of the assignments was relatively consistent between 2015 and 2016 in CE404 and based on the reporting data from CE403, it is likely that a portion of the increased time can be attributed to the use of probabilistic MC and potentially linked to thinking critically about the course material.



Figure 7: Student reported out-of-class time per lesson comparison between 2015 and 2016 CE404 classes. The boxes indicate course assignments: Problem Sets (PS) and Engineering Design Problems (EDP) were out of class work and Written Partial Reviews (WPR) were in-class exams.

Even though a minority of the 2016 CE404 class reported that their critical thought increased as a result of the probabilistic MC questions, all other metrics indicate that the assessments

potentially did affect the students' level of critical thought. As mentioned above the overall student sentiment towards potential losing points on the assessment likely contributed towards some of the lower reports of critical thought.

2. Increases the student's understanding of technical language used in the course

As seen in Figure 8, 45% of the class felt probabilistic MC questions increased their understanding for technical language and only 34% felt the questions did not increase their understanding.



Figure 8: CE404 2016 Student Survey Results for Question 2

3. Increases the student's ability to self-assess their level of understanding

Figure 9 indicates that 42% of the class felt probabilistic MC questions increased their ability to self-assess their own knowledge in the subject while only 26% felt the questions did not increase their ability to self-assess their own knowledge. The majority of the students that left negative feedback in the comments section of the survey form reported frustration that the questions caused them to "second guess" themselves. Another student reported that "engineering problems should have one discrete answer that can be confidently found [making the probabilistic MC questions] just a hassle." The authors were concerned with this response on two levels. The first level was that there is a difference in finding the correct answer and having confidence in finding the correct answer; the later can have significant repercussions if incorrect. Secondly, engineering problems do not always have a discrete correct answer due to varying assumptions, stakeholder values, changing environmental conditions, etc.



Figure 9: CE404 2016 Student Survey Results for Question 3

4. Increases the student's ability to express their level of confidence

While 42% of the class felt probabilistic MC questions increased their ability to self-assess their knowledge, only 18% of the class felt the questions increased their ability to express their level of confidence based on their knowledge state, as seen in Figure 10. Students commented that they resorted to using a high probability for the answer choice they felt was correct regardless of their confidence because they were willing to take the risk in order to get as many points as possible: "my confidence level wasn't even reflected, it was more a gamble" and "I want to get every single point when I get the answer correct."



Figure 10: CE404 2016 Student Survey Results for Question 4

5. Increases the student's self-learning

Figure 11 indicates that only 24% of the class felt probabilistic MC questions increased their self-learning while 45% felt the questions did not increase their self-learning. Figure 12 indicates that the 2016 CE404 students felt that they were more responsible for their own learning (Question A1) and that their classmates contributed to more of their learning (Question A5) than the 2015 CE404 students. The higher response of the 2016 CE404 students to questions A1 and A5 is noteworthy considering that 2016 class reported lower values than the 2015 class on the these questions in CE403. Figure 12 also indicates that the 2016 CE404 students to the 2015 CE404 students, yet the same 2016 students reported both a higher level of responsibility for learning and motivation for learning compared to all other courses in the division, department and academy as seen in Figure 13.



Figure 11: CE404 2016 Student Survey Results for Question 5



Figure 12: CE404 2015 and 2016 end-of-semester feedback comparison



Figure 13: CE404 2016 end-of-semester feedback comparison to Civil Engineering (CE) Division, Civil & Mechanical Engineering (C&ME) Department, and United States Military Academy (USMA) courses

6. Increases the student's quantitative performance on mid-term and end-term exams

The 2016 CE404 class had an average incoming GPA of 0.19 lower than the 2015 CE404 class (3.42 vs. 3.23 on a 4.0 scale), and the class had a 4% lower average on major graded events in CE403. Despite the lower GPA and average grade in structural analysis, the 2016 CE404 class outperformed the 2015 CE404 class by 4.8% across all major graded events as shown in Table 1.

Course Average Change Between 2015 and 2016 (%)					
	Mid-Term 1	Mid-Term 2	Term-End Exam	Overall	
Торіс	(WPRI)	(WPR2)	(TEE)	Overall	
Total	1.92	4.59	(1.69)	4.82	
Conceptual	(5.94)	4.17	1.74	(0.04)	
Tension	1.61		(2.25)	(0.63)	
Compression	5.85		(6.04)	(0.19)	
Flexural		3.27	(2.86)	0.42	
Beam-Column		7.56	2.00	9.56	
Wood Design		(1.69)	(3.43)	(5.12)	

 Table 1: Comparison of 2015 and 2016 CE404 Course Averages on all Major Graded Events by topic

 Course Average
 Change Between 2015 and 2016 (%)

Additionally, the 2016 class increased their average 9.6% in the Beam-Column topic, which is considered the most conceptually difficult topic in CE404 due to the requirement for students to understand 2nd order structural analysis and frame-sway analysis. It is unclear if the use of probabilistic multiple choice can be directly attributed to the overall increased performance.

CONCLUSION

Probabilistic MC was used in a structural design course to examine the impact on student critical thinking, understanding of technical language, self-assessment, confidence, self-learning, and overall performance. While there is no clear indication that probabilistic MC had a significant positive impact in most of the areas, students using the probabilistic MC did score 4.8% higher overall on major graded events, when compared to the previous year when probabilistic MC was not used. The authors feel more research with probabilistic MC should be conducted in technical undergraduate courses before any definite results can be inferred.

BIBLIOGRAPHY

1. Lowman, Joseph. Mastering the Techniques of Teaching. San Francisco : Jossey-Bass, 1995. pp. 254-255.

2. **Patricia Armstrong.** Center for Teaching. *Vanterbilt University*. [Online] July 9, 2015. [Cited: July 9, 2015.] http://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/.

3. *Improving Education through Computer-Based Alternative Assessment Methods*. **Paul, Jody.** Cambridge : Cambridge University Press, 1994.

4. **Priestley, M.** *Performance Assessment in Education and Training: Alternative Techniques.* First. s.l. : Educational Technology Publications, 1982. pp. 179-181.

5. *Different Methods of Multiple-Choice Test: Implications and Design for Further Research.* Ng, Annie W.Y and Chan, Alan H.S. Hong Kong : IMECS, 2009. International MultiConference of Engineers and Computer Scientists. pp. 1958-1963. Vol II.

6. *A Taxonomy of Multiple-Choice Item-Writing Rules*. Haladyna, Thomas M. and Downing, Steven M. 1, s.l. : Lawrence Erlbaum Associates, Inc, 1989, Applied Measurement in Education, Vol. 2, pp. 37-50.

7. Some Modifications of the Multiple-Choice Item. **Dressel, Paul L and Schmid, John.** 4, Durham : Educational and Psychological Measurement, December 1953, Educational and Psychological Measurement, Vol. 13, pp. 574-595.

8. Brown, Thomas A. and Shuford, Emir H. *Quantifying Uncertainty Into Numerical Probabilities for The Reporting of Intelligence*. Defense Advanced Research Projects Agency. Santa Monica : Rand, 1973. pp. 1-9. R-1185-ARPA.

9. Brown, Thomas A. *Probabilistic Forecasts and Reproducing Scoring Systems*. Advanced Research Projects Agency. Santa Montica : Rand, 1970. pp. 1-32. RM-6299-ARPA.

10. *Strictly Proper Scoring Rules in an Absolute Grading*. **Dees, Robert, Gilliam, Ken S. and Kwinn, Michael J.** West Point : American Society of Engineering Education, 2008. 2008 American Society of Engineering Education (ASEE) Zone 1 Regional Conference. 11. Sibley, William L. A Prototype Computer Program for Interactive Computer-Administered Admissible Probability Measurement. Rand. Santa Monica : Defense Advanced Research Agency, 1974.

12. A Scoring Method Based on Simple Probability Theory that Considers Partial Knowledge and Omission of Answers in Multiple-Choice Testing. Adair, Desmond and Jaeger, Martin. 2013, International Journal of Engineering Education, pp. 974-985.

13. *Confidence Testing: A New Tool For Measurement*. **Shuford, Emir H.** Governors Island : Air Force Office of Scientific Research, 1969. 11th Annual Conference of the Military Testing Association. pp. 1-13. AFOSR 69-2348TR.

14. *Confidence Assessment in the Teaching of Basic Science*. **Gardner-Medwin, A R.** s.l. : Association for Learning Technology Journal, 1995, Vol. 3, pp. 80-85.

15. *The Use of Confidence Testing in the Academic Instructor Course.* Gardner, Willie C. New York : Air Force Office of Scientific Research, 1961. Annual Conference of the Military Testing Association (11th, Governors Island, New York, September 15-19, 1969). AD-703-726.

ENCLOSURE 1



Figure 13: CE403 end-of-semester feedback comparison between CE404's Class of 2016 and 2015

CE403 Course Averages Comparison					
Graduating Year Group	Major Graded Event	Out-of-class Time / Lesson (min)			
2016	80.7%	76			
2015	84.7%	85.8			

-4.03%

-9.80

Difference

Table 2: Comparison of 2015 and 2016 CE404 Student Averages in CE403