Mechanics Test Item Analysis and Focussed Feedback

Peter Rosati
The University of Western Ontario

Abstract

Multiple-choice questions although not recommended as the only form of course assessment are a useful and efficient component for evaluating large enrolment Mechanics classes and for comparing, and if necessary standardizing, performance between parallel class sections.

The results from seven multiple-choice examinations in two sections of a Statics course show that the average aggregate examination mark difference between the two class sections was only 2.5% and that these differences were not significant in any single examination: the students had learned Statics to approximately the same level in each class section.
The average student performance on individual multiple-choice questions differed more widely between the two class sections but even so this performance difference was significant for only two of the ninety-four multiple-choice questions: probably representing extra emphasis or specific practice on those particular sub-topics in one of the class sections.

An analysis of the group of multiple-choice questions on which the students performed well in comparison with those questions on which they performed poorly provided a clear indication of those topics - overall concepts, geometrical items, position of resultant force, moment equations, friction - that the students find difficult and consequently need teaching emphasis in future courses.

An experiment involving the reissue in a later test of slightly altered items from a previous test showed that quite subtle differences in item presentation can have a dramatic effect on student performance for these items.

Introduction

In the large engineering undergraduate classes it becomes more attractive to design and evaluate, at least partially, using machine scored tests. Such tests, although more time consuming to design, are quickly graded by computer which also provides item performance statistics that contain useful detailed information for the course instructor.

Some might argue that the learning in a problem-solving course such as Mechanics cannot be adequately assessed by multiple-choice questions. However, Hudson[1] has shown that for over 90% of students M-C.Q tests gave approximately the same grade as hand graded tests. Rosati[2] and Pfeiffenberger[3] analysed the performance in Mechanics tests of engineering and physics students respectively, and found high correlations between M-C.Q scores and free response scores. Of course it is necessary to use several M-C.Q's on the one topic to adequately assess that topic and a test consisting of "more M-C.Q's means less weight for each with an effect similar to the awarding of partial credit"[4] in grading free response problems. One of the instructional goals of most engineering courses is that students also develop the habit of professionally formatting their problem solutions. This style of presentation cannot be tested on an M-C.Q test so must either be assessed in homework assignments or in a hybrid test that contains some M-C.Q test items and some free response problems.

The expert/novice problem-solving literature gives an insight into which test items will likely be difficult for students and which will be easy. Larkin[5] found that experts worked forward through the problem and that novices work backwards from the unknowns. The expert tended to explore the problem qualitatively before solving whereas the novice plunged immediately into a mathematical description or formula. Clements[6] researched the problem-solving processes of engineering students and found that many of them, although able to successfully solve a formula centered problem, did not understand the conceptual model or physical situation appropriate to the problem. Zajchowski[7] observed strong students and weaker students as they solved mechanics problems and found that the strongest students organized their knowledge on the underlying principles whereas weaker students organized around memorized formulae and isolated problem features. Woods[8] provides a good summary of the expert/novice problem-solving work of many researchers and claims that successful problem-solvers focus on "actions" (rather than "objects") connect to worked examples that have the same processes and principles (rather than look alike or have the same objects) verbally emphasize key words and disregard irrelevant words, can identify the information necessary to solve (rather than using all the given information) and can draw detailed explicit diagrams.

This paper presents the results of ten M-C.Q tests in Statics and compares student performance both for two different lecture sections and also for each multiple-choice item. A detailed analysis of the test items provides useful insight regarding the characteristics of easy and difficult items and hence the probable basis of the students' difficulties in Statics.

Results and Discussion

Overall Performance on Multiple-choice Tests

At the University of Western Ontario, Statics is taught in two sections with about 75 students in each. Both sections write common examations which consists of M-C.Q's (40%) and free response problems (60%). The aggregate performance on 130 M-C.Q's from 10 of these examinations (1993-95) is shown in Figure 1. M-C.Q's which more than 70% of students answered correctly were termed "easy" and M-C.Q's which less than 30% of students answered correctly were termed "difficult". Characteristics of the M-C.Q's from each of these two categories are discussed below.

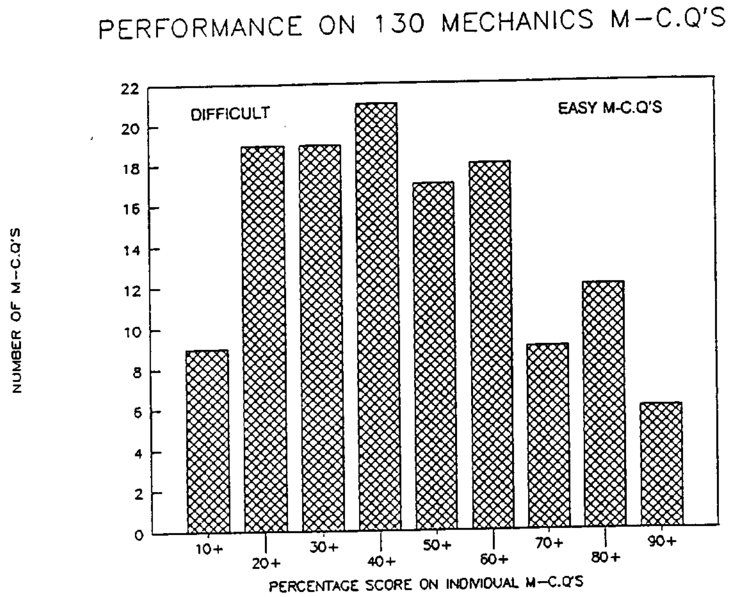


Figure 1 : **Overall student performance on 130 multiple-choice questions.**

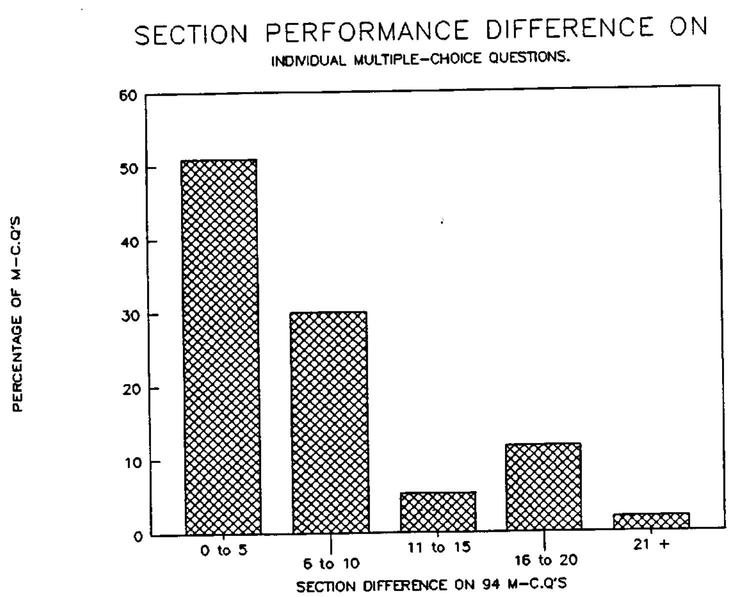


Figure 2 : **Difference in section performance on 94 multiple-choice questions.**

The difference between the average scores of the two class sections is shown in Figure 2. Fifty percent of the questions show a section difference below 5% and 80% of the questions show a section difference below 10%. An analysis of those multiple-choice questions which produced an average section performance difference above 15% did not reveal them to be on particular topics nor to be of greater nor lesser difficulty than the other M-C.Q's. For only two of the individual questions was the performance difference between sections statistically significant at $p < 0.001$ on a chi-square test. These significant performance differences on the two items are probably attributable to the class instructors in the two lecture sections inadvertently placing different emphases on specific topics and particular practise example problems which favoured a superior performance on these two M-C.Q's. However, the average difference in M-C.Q scores between the two section on any given examination was only 2.5% and these average section differences were not significant for any of the individual examinations.

Easy and Difficult Test Items

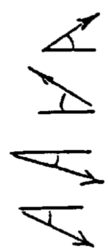
The students' performance on 130 M-C.Q's is shown at Figure 1. Those M-C.Q's for which more than 70% of the students answered correctly were categorised as the 'easy' questions and they form the top 21% of the 130 M-C.Q's. Conversely, those M-C.Q's for which less than 30% of the students answered correctly were termed 'difficult' questions and they account for the bottom 21% of the total 130 M-C.Q's.

Characteristics of the 'easy' M-C.Q's were that they were mostly straightforward algorithms; to express a direction as a unit vector, to use the dot product to solve the angle between two vectors, to evaluate the friction forces on a sliding block or the magnitude of the resultant of a given distributed load diagram or of a given planar force system, to repeat a memorised definition such as a cross product or a moment or to solve a straightforward 2-D equilibrium problem. An example of a straightforward 2-D equilibrium is shown in Figure 3. The information is presented numerically and one application of the moment equilibrium equation about point C produces the required numerical answer.

The main characteristics of the 'difficult' M-C.Q's were that they tested more the overall concepts, involved geometry, needed the application of the moment equation to establish equilibrium or the position of the resultant force or demanded an understanding of the overall behaviour of a truss under load or were friction problems that required judgement to determine either the direction of the friction force or whether the friction force had reached its limiting value. A typical example of a difficult problem is shown at Figure 4. It is a 2-D equilibrium problem presented in algebraic form and, although it can be solved (somewhat long-windedly for a multiple-choice question!) from the three scalar equations of 2-D equilibrium, it can most readily be solved by a student with a sure grasp of resultants and the equilibrium of a three force body.

The uniform beam AB weighs W and also carries a load W at B. The boom is supported by a cable DC attached at the mid-point C. The direction of the reaction at A on AB is:

a). horizontal



b). 30°

c). 30°

d) 11°

e). 30°

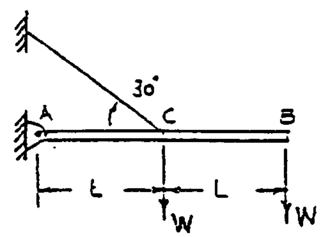


Figure 4 : Example of a Difficult Test Item.

Examples of problems that exposed weakness in geometry were particle equilibrium problems where the angles needed to be deduced, or force component resolutions along non-rectilinear

The reaction at C is:

a). 50 N ↑

b). 50 N ↓

c). 67 N ↑

d). 100 N ↑

e). 150 N ↑

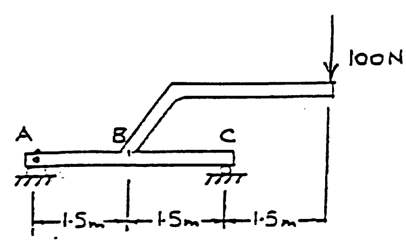


Figure 3 : Example of an Easy Test Item.

axes, or method-of-section truss problems in which moment arm dimensions needed to be calculated or three-force rigid body equilibrium problems requiring the application of a vector triangle solution to provide the force directions.

Focussing Test Items for Insights

As an experiment to probe the precise difficulties that students were finding with the M-C.Q's, some of the easy items were altered, quite slightly, in a way assumed to make them more difficult and reprocessed in this new form on a later examination. For example, a particle resultant problem with zero resultant horizontal component was reissued as a particle problem with a resultant horizontal component of 10 lb. Student success on the problem dropped from 88% to 66%. An item about the resultant of two vectors where the diagram cued the appropriate parallelogram of forces (93%) was recycled as a vector addition not cued by the diagram (60%). Approximately 96% of students solved the magnitude of the resultant of a coplanar force system but only 51% correctly positioned the same resultant. A block sliding on an inclined plane can be correctly solved for the angle causing slip (88%) but apparently becomes much more difficult if equilibrium is destroyed by tipping (16%). The vertical component of a simple beam reaction (72%) is easy but the total reaction if the roller support is on an inclined plane (25%) is much more difficult.

The same experiment can be tried in reverse: to slightly simplify a difficult question into a simpler one. Care is needed, in this case, not to oversimplify the question nor to "give the game away" with a clue that is too overt. For a planar force system acting on a column only 23% of students could reduce the system to a resultant but 66% correctly reduced the same system to a force and couple at the column base. Resolving a force along non-rectilinear axes was only 19% successful but recycled with the hint "draw the vector triangle" produced 57% correct. An item to obtain the bending moment at a point in a beam from a given shear force diagram (6%) was improved (30%) but not as much as anticipated by the hint "the change in bending moment is the area under the shear force diagram". Students solving for the number of compression members in a truss (27%) were not helped much (35%) by the hint "consider the effect on the truss of removing the member in question".

The well known friction problem involving a rectangular block on a rough horizontal surface subject to a horizontal force and the question "will the block slip or tip" became an 'easy' question (91%) for the students if the correct answer was designed to be "slip" and became a 'difficult' question (19%) if the correct answer was designed to be "tip". In other words, the weaker students immediately flagged such a problem as a friction problem, calculated the friction force at slip by plugging into $F = N$ and selected the slipping distractor. They did not consider the overall equilibrium of the block and the possibility that it might tip.

Lessons learned from these experiments in fine-tuning the test items are that quite subtle differences in item presentation can have a dramatic effect on student performance. In general, though, there were few surprises. The problem-solving literature (and also experienced engineering educators) predicted that many students can use equations and superficially get correct answers to algorithmic problems without understanding the underlying concepts. This analysis of M-C.Q items and the evidence of the 'easy' and 'difficulty' M-C.Q's give strong support that many students can solve well-practised routines and algorithms but still have considerable difficulty with geometry, overall concepts, resultant or equilibrium problems involving moments and friction problems requiring judgement. The implication for engineering educators is for them to take steps to both teach and to test the concepts and underlying physics as well as the skills and algorithms.

Conclusion

Multiple-choice questions are a useful and efficient component for testing large enrolment Mechanics classes and for comparing, and if necessary standardising, performance between parallel student sections.

The item analysis over several M-C.Q examinations pinpoints for the instructor those aspects of the course topics that the students have had difficulty with and suggests those specific topics which might need extra emphasis in the teaching of future courses.

References

1. Hudson, H.T. and Hudson, C.K. " Suggestions on the construction of multiple-choice tests", American Journal of Physics, Vol.40, No.9, 1981, pp.838-841.

2. Rosati, P.A., "Using multiple-choice questions in Mechanics", Frontiers in Education Conference Proceedings, 1987, pp.248-253.

3. Pfeiffenberger, G.W. and Modn, C.C., "A validity study of the multiple-choice component of the advanced placement Physics C examination", American Journal of Physics, Vol.45, No.11, 1977, pp.1066-1069.

4. Leuba, R.J., "Machine-scored Testing: Purposes, principles and practice", Engineering Education, Vol.77, No.2, 1986, pp.89-95.

5. Larkin, J.H., McDermott, J., Simon, D.P. and Simon, H.A., "Expert and novice performance in solving physics problems", Science, Vol.208, 1980, pp.335-1342.

6. Clements, J., "Solving problems with formulas: some limitations", Engineering Education, Vol.72, No.2, 1981, pp.158-162.

7. Zajchowski, R. and Martin, J., "Differences in the problem-solving of stronger and weaker novices in Physics: knowledge, strategies or knowledge structure", A.E.R.A., Chicago, 1991.

8. Woods, D.R., "PS News", No.55, 1988, pp.2-18.

PETER ROSATI is a Professor in the Department of Civil Engineering at The University of Western Ontario. He has engineering degrees from Oxford and Western, and an education degree from West Virginia University. His research in engineering education has focussed on problems associated with personalizing the large enrolment mechanics courses. He has implemented a successful Keller plan course in dynamics, designed computer problem-solving routines and is currently principal investigator in a seven-year longitudinal study relating engineering student performance and learning style to the MBTI personality type.