EFFECT OF MULTIPLE CHOICE QUESTION FORMAT ON STUDENT PERFORMANCE

Umairia Malik

Umairia Malik (u.malik@student.adfa.edu.au)
School of Science, University of New South Wales Canberra, Canberra, ACT 2610, Australia

KEYWORDS: gender gap, question format

ABSTRACT

As we move to online teaching and assessment, it is timely to consider how multiple-choice quizzes can best be used to give accurate information about student knowledge and learning. The gender gap in performance often has been observed on multiple choice quizzes, including commonly used physics inventories such as the Force Concept Inventory (FCI). Many factors have been identified that contribute to these gaps, however the gaps have not been closed. In this paper, the effect of question format of the six Identifying Force (IF) questions of the FCI was investigated. These questions are presented as Complex Multiple Choice (CMC) questions and were modified to Multiple True-False (MTF) format. Comparing the results of the original (CMC) and modified (MTF) question format, I found that both male and female students' performance was lower when MTF format was used and the gender gap was not reduced. A comparison of students' answer choices on the two formats indicated that the CMC (original) format overestimated students' knowledge and failed to identify some misconceptions. Hence, although modifying the format from CMC to MTF did not reduce the gender gap, it did provide more information about students' thinking.

Proceedings of the Australian Conference on Science and Mathematics Education, 30 September - 2 October 2020, pages 123-128, ISBN Number 978-0-9871834-9-1.

INTRODUCTION

With so many of our classes and assessments moved online in 2020 in response to the COVID-19 crisis, and many people experimenting with different online assessment types, it is timely to consider the effect of question format. While Learning Management Systems (LMSs) allow many different types of questions, multiple-choice questions (MCQs) are particularly easy to implement and allow for automated marking (Aiken, 1982). Hence, they are a tempting option for instructors under pressure to quickly move their assessment online.

However, the multiple-choice questions (MCQs) have received some warranted criticisms. While they may allow quick assessment of students' declarative and, if well written, their conceptual understanding (Bates, Donnelly, MacPhee, Sands, Birch & Walet, 2013; Henderson, Stewart, Stewart, Michaluk & Traxler, 2017; Henderson, Stewart & Traxler, 2019), they are less effective at assessing procedural knowledge. It is difficult to assess skills such as problem solving using MCQs, particularly in physics where the drawing of a diagram is often part of the solution process. In addition, it has been demonstrated that MCQ assessments tend to favour male students over female students because of the different ways males and females approach questions and the different strategies they tend to use (Stenlund, Eklöf & Lyrén, 2017). Gender gaps have been commonly observed (Madsen, McKagan & Sayre, 2013; McCullough, 2004; Wilson, Low, Verdon & Verdon, 2016) in standard diagnostic tests such as the Force Concept Inventory (FCI) (Halloun, Hake, Mosca, & Hestenes, 1995). Therefore, the choice to use MCQs should be made with caution, and with attention paid to what exactly is being assessed, and consideration of possible gender bias.

If the decision to use MCQs is made, then there is a choice of MCQ formats that can be used. When we think of an MCQ, we usually think of a question stem, followed by four or five answer options, of which only one is the correct answer. This is a simple multiple-choice (SMC) question. It is also possible, and easy to implement in Learning Management Systems (LMSs), SMCs where multiple answer options are correct. Another format of an MCQ is the complex multiple-choice (CMC) format. In CMC format the question consists of a stem with a list of typically three or four items/statements, and answer options which are different combinations of the listed statements. An example of a CMC format question is given in Figure 1a below. Students choose only one answer option, and these CMC questions are marked as typical MCQs (Ma, 2004). This format was particularly useful before the widespread use of online quizzes, because students can answer them on the same generic "optical mark recognition" as SMC papers, allowing automatic marking. A limitation of CMCs is that with three

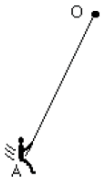
or four items in the list, only a small fraction of the possible combinations of listed items can be made available as answer options (Haladyna & Downing, 1989). Hence students may not be able to choose an answer corresponding to their understanding or may have a better chance of guessing correctly using strategies such as elimination of incorrect answers.

A format which overcomes this limitation is the multiple true-false (MTF) format. The MTF format consists of a stem followed by a list of sub-questions, as shown in Figure 1b. The sub-questions are analogous to the list of items/statements in the CMC format, however a student must respond to each sub-question independently as either true or false, and so every possible combination of items/statements is available as an answer option (Cronbach, 1939; Frisbie & Sweeny, 1982). Hence, in contrast to typical MC format, the MTF format generates multiple responses through which students' thinking can be thoroughly examined. Results of a recent study (Brassil & Couch, 2019) have shown that students' correct response rate in SMC format was higher than the MTF format, suggesting that question format influences students' performance. With online implementation, this format is just as easy to use and automatically mark as SMC or CMC questions.

(a)

The figure below shows a boy swinging on a rope, starting at a point a higher than A. Consider the following distinct forces:

1. A downward force of gravity.
2. A force exerted by the rope pointing from A to O.
3. A force in the direction of boy's motion.
4. A force pointing from O to A.



Which of the above forces is (are) acting on the boy when he is at position A?

A. 1 only.
B. 1 and 2.
C. 1 and 3.
D. 1, 2 and 3.
E. 1, 3, and 4.

(b)

The figure below shows a boy swinging on a rope, starting at a point higher than A. Which of the following forces is (are) acting on the boy when he is at position A?

1. A downward force of gravity.	Y	N
2. A force exerted by the rope pointing from A to O.	Y	N
3. A force in the direction of the boy's motion.	Y	N
4. A force pointing from O to A.	Y	N

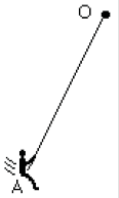


Figure 1. (a) Question 18 of the FCI in CMC format. (b) Question 18 of the FCI in MTF format

The Force Concept Inventory (FCI) (Halloun, Hake, Mosca, & Hestenes, 1995) uses a combination of SMCs and CMCs to assess students' Newtonian thinking (Hake, 1998; Bates, Donnelly, MacPhee, Sands, Birch, & Walet, 2013). Of the 30 MCQs of this inventory, 24 questions have typical SMC format and six questions have the CMC format. All six of the CMC format questions are Identifying Force (IF) questions in which a list of forces is given, and the answer options are different combinations of these forces (questions 5,11, 13, 18, 29 and 30). As with the rest of the FCI, these questions show a gender gap that favours males (Osborn Popp, Meltzer, & Megowan-Romanowicz, 2011).

This study investigates how changing the question format from CMC to MTF for the six Identifying Force (IF) questions of the FCI affects performance of male and female students.

METHOD

The University of New South Wales (UNSW) Sydney provides an introductory calculus-based physics course called Physics 1A. In 2018, this course ran over twelve weeks in each of two teaching semesters. In 2019 the university's teaching mode shifted to a three-term system, with each term consisting of a 10-week teaching period. Physics 1A, which begins with mechanics, was taught in both semesters, and during all terms.

The data was gathered from the second semester (S2, n = 207, with 72% males and 28% females) in 2018 and the second and third terms (T2 and T3, with a total of n = 542, 65% males, 35% females) in 2019. Approximately eighty percent of the students were engineering students in all cohorts. Students who participated in this study were at approximately the same stage of their degree – part way through their first year, and about to commence their first university level physics course.

In 2018 the S2 Physics 1A cohort was given the FCI in its original format – a combination of SMC and CMC questions. In 2019 the T2 and T3 Physics 1A cohorts were given a modified version of the FCI such that the CMC questions were replaced with MTF questions. In each case, the question stem was unchanged, and the list of forces was given as a set of sub-questions, as shown in Figure 1b above. In 2018 the 30 questions of the FCI were marked using a number right scoring method in which one mark was given for the correct answer and zero for an incorrect answer or unanswered question to give a score out of 30. This was subsequently converted to a percentage. In 2019, in order to mark the six IF questions given in the MTF format, one mark was awarded only for a *fully* correct answer and zero otherwise, to enable scores to be compared across the cohorts. For example, in Figure 1a above, CMC format, the correct answer is B and a student choosing this answer is awarded one mark. In Figure 1b above, MTF format, the fully correct answer is 'yes' to the sub-questions 1 and 2 and 'no' to the sub-questions 3 and 4; therefore, this combination of answers is awarded one mark, and any other combination is awarded zero marks. The combination of answers in Figure 1b is equivalent to answer choice B in Figure 1a.

A one-way Analysis of Variance (ANOVA) was used to investigate the similarity of the three datasets on the unchanged (SMC) questions. The significance of differences between males' and females' scores was investigated using *t*-tests.

RESULTS AND DISCUSSION

The unchanged (SMC) questions provide a check that any change in performance on the modified questions is indeed due to the changed format, and not a difference in the cohorts of students. Table 1 below shows the average percent scores (standard error) of male and female students on the unaltered questions for the three cohorts.

Table 1: Average percent scores (standard error) of males and females, and *p*-values in 2018 and 2019 on the 24 unchanged questions

Years	2018-S2	2019-T2	2019-T3	*P*-values
Total males/females	149 / 58	251 / 110	104 / 77	-
Males Avg (s.e)	67.2 (1.8)	66.3 (1.2)	65.7 (2.0)	0.84
Females Avg (s.e)	54.2 (2.8)	51.5 (2.0)	51.8 (2.2)	0.72

Based on an Analysis of Variance (ANOVA), the difference in the average percentage scores for the three datasets on the 24 unchanged questions was not statistically significant for either males or females (*p*-values shown in Table 1). This gives us confidence in ascribing any change in performance on the six modified IF questions to differences in question format. In addition, it allows us to merge the datasets of the two terms of 2019, referred to simply as 2019 hereafter.

The results for the six modified questions are presented in Table 2. These results clearly indicate that changing the format of the IF questions significantly decreased the average scores for both genders but did not affect the gender gaps.

Table 2: Average percent scores (standard error) of males and females, and *p*-values in CMC format (2018) and MTF format (2019) on the six identifying force questions

	Formats	
	CMC	**MTF**
Total males / females	149 / 58	355 / 187
Males Avg (s.e)	56.5 (2.8)	44.6 (1.7)
Females Avg (s.e)	42.2 (4.2)	32.2 (2.2)
Gender gaps (s.e)	14.2 (5.1)	12.4 (2.8)
P-values	0.003	6.68×10^{-06}

As there was no change in scores on the unaltered FCI questions, the decrease in scores on the modified questions indicates that the CMC format was overestimating students' knowledge. The correct answer in CMC format was selected by some students, who would not have done so in MTF format. Hence the CMC format does not give as accurate a measure of students' knowledge as the MTF format does.

To explore in detail how the change in format affected students' responses, a question by question analysis was performed. The analysis for one of the questions is presented here as an example.

AN EXAMPLE: FORCES IN CURCULAR MOTION

In question 18 (Figure 1a above) of the FCI, students were asked which forces are acting on a boy moving in circular motion. There are five possible answers to this question in CMC format, however, in the MTF format, Figure 1b, there are 16 possible answers. To allow comparison of students' answer choices, students' responses on the MTF format (Figure 1b) were categorised as either equivalent to one of the CMC answer choices (A, B, C, D or E) or 'Other'. For instance, if a student responds 'yes' to sub-questions 1 and 2 and 'no' to sub-questions 3 and 4, this is equivalent to the correct answer choice B in the CMC format. If a student selects option B, this shows that they are correctly recognising that there are only two forces acting on the boy – a force of gravity and a tension force exerted by the rope from A to O. Responding 'yes' to sub-questions 1,2, and 3 and 'no' to 4 in the MTF format is equivalent to answer choice D in the CMC format. This answer choice identifies a common misconception held by students – that there is always a force in the direction of motion. Any student responses on the MTF format which were not equivalent to CMC answer choices were grouped in the 'Other' category.

Figure 2 below shows the fraction of students choosing each answer option for the two versions of the question. Comparing Figures 2a and 2b below, we can see that the change in format resulted in a decrease in correct response rate. We can also see that for both formats, more males answered correctly than females and there was a significant gender gap, although this decreased with the change in format. However, the decrease in the gender gap from 13% in the CMC format to 9% in the MTF format was not significant (*p*-value = 0.64).

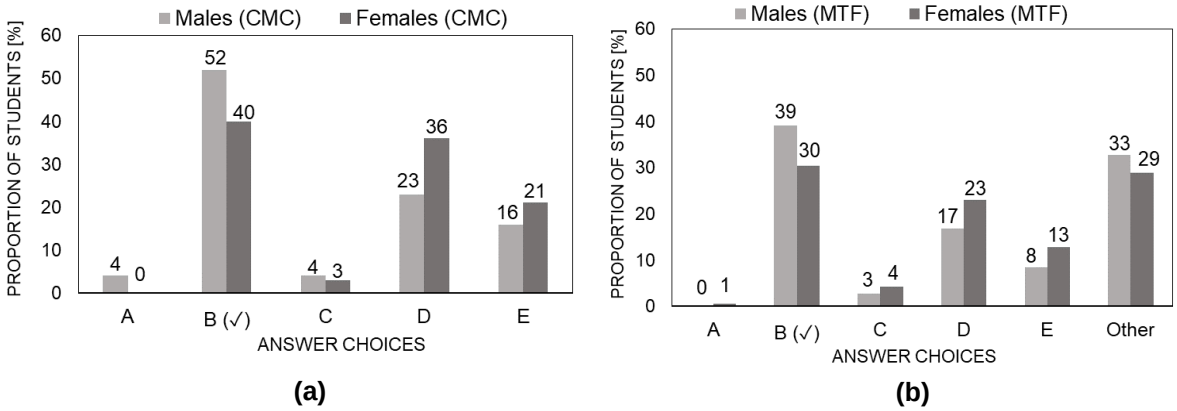


Figure 2: Percentage of males and females choosing answer choices on question 18 of the FCI in CMC format Figure 2a (left), and in MTF format in Figure 2b (right).

If a student responds ‘yes’ to the sub-question 3 in the MTF format or chooses any of C, D or E in the CMC format, this indicates a belief in a force in the direction of its motion. Responding ‘yes’ to the sub-question 4 or answer E in the CMC format shows that a student has the misconception that there is a centrifugal force (see Eaton, Vavruska, & Willoughby, 2019 for an updated FCI distractor taxonomy for 1995 version). In the MTF format we can clearly identify which individual misconceptions a student has – or, what fraction of a class holds that misconception. However, with the limited answer options available in the CMC format we cannot.

The magnitude of the information lost by the CMC format about students’ knowledge is shown by the size of the ‘Other’ columns in Figure 2b. Almost one third of students have an understanding which is not captured by the CMC format, however, it is captured by the MTF format.

When we compare Figures 2a and 2b we can see which options from the CMC students chose when they could not choose exactly the combination of forces they believed to be acting. From CMC to MTF, the percentage of students selecting options B, D and E all decreased. The drop in frequency of correct answer choice (B) by approximately 10% for both genders indicates that a substantial number of students do not have the level of understanding of forces indicated by the CMC format of the question. These students have chosen the correct answer for the wrong reasons – in some cases a lucky guess as there is a 20% chance of guessing correctly on the CMC format compared to only a 6% chance on the MTF format. In other cases, a false positive may be due to strategic elimination of incorrect answers.

Similar to this example, the results of all the other five CMC (also IF) questions showed that the correct responses decreased for both male and female students when they attempted the MTF format. However, the gender gaps varied from question to question.

CONCLUSION AND IMPLICATIONS

This study has shown that changing question format from CMC to MTF significantly decreased performance of both genders on the six IF questions of the FCI. This shows that the CMC format overestimates students’ knowledge compared to the MTF format. This could be due to the CMC format providing clues to students by limiting the number of answer choices. For instance, in question 18 of this study, the four forces given in the stem would result in 16 possible answer choices in the MTF format. Furthermore, identifying correctly that a force is not acting on the object would help students to eliminate answer choices which contains that force, and, as a result, a student’s chances to answer correctly increases. In addition, the chances of a correct guess are much higher on the CMC format questions.

In the example shown, almost one third of students chose combinations of forces which were not captured in the CMC format. This indicates that not only almost one third of the information about student understanding is lost by using CMC format rather than MTF, but also students’ marks do not necessarily represent their understanding. Through the MTF format, instructors can gain more information about students’ thinking and modify their instruction to improve students’ learning.

While it was hoped that changing the question format would reduce the gender gap, this was not the case. Although the MTF format did not significantly reduce the gender gap, neither did it increase it – the change at least did not exacerbate the gap.

In conclusion, in these unprecedented times, when moving assessment online if instructors choose to use MCQs, they should be aware that there are a range of formats possible and consider carefully their choice of question format.

ACKNOWLEDGEMENT

The author gratefully acknowledges the receipt of an HDR completion scholarship provided by Commonwealth Australia, Elizabeth Angstmann for sharing data used in this study, and supervisors for providing continuous support and valuable feedback on this research work.

REFERENCES

Aiken, L. R. (1987). Testing with multiple-choice items. *Journal of Research & Development in Education*, *19*(1), 27–31.

Bates, S., Donnelly, R., MacPhee, C., Sands, D., Birch, M., & Walet, N. R. (2013). Gender differences in conceptual understanding of Newtonian mechanics: a UK cross-institution comparison. *European Journal of Physics*, *34*(2), 421-434. doi:10.1088/0143-0807/34/2/421.

Brassil, C., & Couch, B. (2019). Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: a Bayesian item response model comparison. *International Journal of STEM Education, 6*(1), 1-17. doi:10.1186/s40594-019-0169-0.

Cronbach, L. J. (1939). Note on the multiple true-false test exercise. *Journal of Educational Psychology*, *30*(8), 628-631. doi:10.1037/h0058247.

Eaton, P., Vavruska, K., & Willoughby, S. (2019). Exploring the preinstruction and postinstruction non-Newtonian world views as measured by the Force Concept Inventory. *Physical Review Physics Education Research*, *15*(1), 010123. doi:10.1103/PhysRevPhysEducRes.15.010123.

Frisbie, D. A., & Sweeney, D. C. (1982). The relative merits of multiple true-false achievement test. *Journal of Educational Measurement*, *19*(1), 29-35. doi:10.1111/j.1745-3984.1982.tb00112.x.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64-74. doi:https://doi.org/10.1119/1.18809.

Haladyna, T. M., & Downing, S. M. (1989). Validity of a Taxonomy of Multiple-Choice Item-Writing Rules. *Applied Measurement in Education*, *2*(1), 51-78. doi:10.1207/s15324818ame0201_4.

Halloun, I., Hake, R., Mosca, E. & Hestenes, D. (1995). Force Concept Inventory. Retrieved from http://modeling.asu.edu/R&E/Research.html (password protected).

Henderson, R., Stewart, G., Stewart, J., Michaluk, L., & Traxler, A. (2017). Exploring the gender gap in the conceptual survey of electricity and magnetism. *Physical Review Physics Education Research*, *13*(2), 020114. doi:10.1103/PhysRevPhysEducRes.13.020114.

Henderson, R., Stewart, J., & Traxler, A. (2019). Partitioning the gender gap in physics conceptual inventories: Force Concept Inventory, Force and Motion Conceptual Evaluation, and Conceptual Survey of Electricity and Magnetism. *Physical Review Physics Education Research*, *15*(1), 010131. doi:10.1103/PhysRevPhysEducRes.15.010131.

Ma, X. (2004). An investigation of alternative approaches to scoring multiple response items on a certification exam (Doctoral dissertation). University of Massachusetts Amherst.

Madsen, A., McKagan, S. B., & Sayre, E. C. (2013). Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap? *Physical Review Special Topics-Physics Education Research, 9*(2), 020121.

McCullough, L. (2004). Gender, context, and physics assessment, *Journal of International Women's Studies*, *5*(4), 20-30. Available at http://vc.bridgew.edu/jiws/vol5/iss4/2.

Osborn Popp, S.E., Meltzer, D.E. & Megowan-Romanowicz, C. (2011). Is the Force Concept Inventory biased? Investigating differential item functioning on a test of conceptual learning in physics. Paper presented at 2011 Annual Meeting of the American Educational Research Association. New Orleans, Louisiana: http://www.aera.net/repository.

Stenlund, T., Eklöf, H., & Lyrén, P.-E. (2017). Group differences in test-taking behaviour: an example from a high-stakes testing program. *Assessment in Education: Principles, Policy & Practice*, *24*(1), 4-20. doi:10.1080/0969594X.2016.1142935.

Wilson, K., Low, D., Verdon, M., & Verdon, A. (2016). Differences in gender performance on competitive physics selection tests. *Physical Review Physics Education Research, 12*(2), 020111. doi.org/10.1103/PhysRevPhysEducRes.12.020111.