Learning Vectors Online: Comparing Multiple-Choice to Drawing Vectors

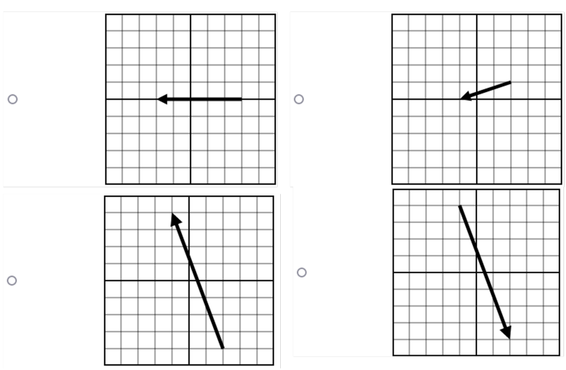
M. Kline[1]

[1]The Ohio State University, Columbus, OH 43210, USA

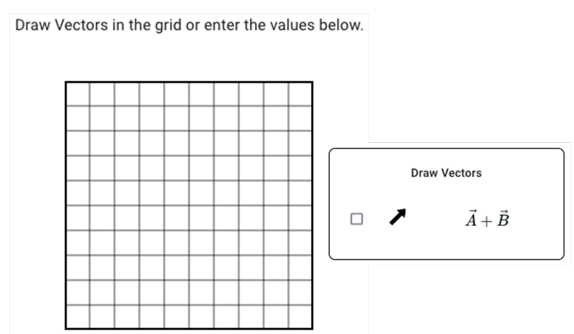
November 11, 2022

1 Introduction

It is well established that introductory physics students have difficulties with vector addition and subtraction, particularly vectors on a grid [1–10]. Heckler and Scaife [2] investigated the *ijk* answer format and the arrows on a grid answer format for vector addition and subtraction and found that students have trouble with the arrow representation itself. We further this investigation into students' difficulties with vector addition by examining two answer formats using the arrow representation: multiple-choice (MC) and free response (FR). Examples of the two answer formats can be found in Fig. 1.



(a) Multiple-Choice



(b) Free Response

Figure 1: Examples of multiple-choice and free response answer formats. In the multiple-choice format, students choose the correct answer from a list of options. In the free response format, students draw the answer vector on a grid.

For more complex physics questions separate from vector addition and subtraction, the typical trade-off between multiple choice and free response answer formats is that the free response format is more insightful and allows for partial credit, but the multiple-choice format is much easier to grade. However, some studies [11, 12] have shown that carefully structured and graded multiple-choice assessments can be viable replacements for assessments in the free response format.

Lastly, the constructed figural response answer format, in contexts outside of vector addition and subtraction, was shown to be more difficult, slightly more discriminating and reliable, and had higher omit rates than multiple choice [13]. The constructed figural response format is similar to our free response format in that the formats require constructed responses, as opposed to choosing from a list of options as with the multiple-choice format.

This study contrasts the multiple-choice and free response item formats for vector addition. First, we investigate their utility in assessing students' understanding of vector addition, then we explore what benefits there are to building fluency through training in the two formats.

2 Participants and Methods

The participants in this study were students enrolled in the first semester (Mechanics) introductory physics course at The Ohio State University. Participants included those enrolled in the algebra-based version of the course, primarily designed for pre-medical students, and the calculus-based version, primarily designed for engineering students.

All students took an online pre-test at the beginning of the term where they were given full credit for completing the assignment. Students were assigned to either the FR training condition or the MC training condition, where the students practiced vector addition throughout the term in online mastery-based assignments only in their designated format. At the end of the semester, all students then took the post-test, which was identical to the pre-test, and again were given full credit for completion.

In the spring term of 2022 (SP22), the algebra-based course had four lecture sections, where two sections were assigned to the MC training condition ($N = 137$) and the other two were assigned to the FR training condition ($N = 176$). The calculus-based course had three sections, where one section was assigned to the MC training condition ($N = 84$) and the other two were assigned to the FR training condition ($N = 235$). The pre- and post-tests for the SP22 term consisted of five vector addition questions in each format. The questions were chosen in isomorphic pairs containing one question in the MC format and one question in the FR format.

This experiment is being repeated in the autumn term of 2022 (AU22) with a few differences. Instead of sections being assigned to a training condition, all students are randomly assigned to a training condition. We also included a control group that did not practice vector addition in either format and instead practiced other content. The pre- and post-tests for the AU22 term consisted of 3 vector addition, 3 vector subtraction, and 2 vector components questions in each format, again in isomorphic pairs. We will discuss results from the pre-test only, as training is still occurring at the time of writing this paper.

In the AU22 term, the algebra-based course had 231 students in the control group, 247 students in the MC training condition, and 237 students in the FR training condition. The calculus-based course had 316 students in the control group, 309 students in the MC training condition, and 301 students in the FR training condition.

Students that did not consent to their data being used in this research were removed from the data set. Additionally, students that did not complete the pre-test or the post-test were also removed. Lastly, students that answered quickly on more than 20% of the questions were removed, as this behavior was associated with rapid guessing and not effort towards answering the questions [14].

3 Results

3.1 Assessment

3.1.1 Item Difficulty

Across both terms and versions of the course, and for both the pre- and post-tests, the free response format was more difficult than the multiple-choice format (Table 1). This difference in difficulty could be due to guessing behavior and the process of eliminating answers they know are incorrect. In the multiple-choice format, there were between 2 and 5 answer choices, so if a student guesses the answer blindly, they will correctly guess the answer 20-50% of the time. They can also improve this accuracy if they can rule out one or more items they know are incorrect.

3.1.2 Item Response Times

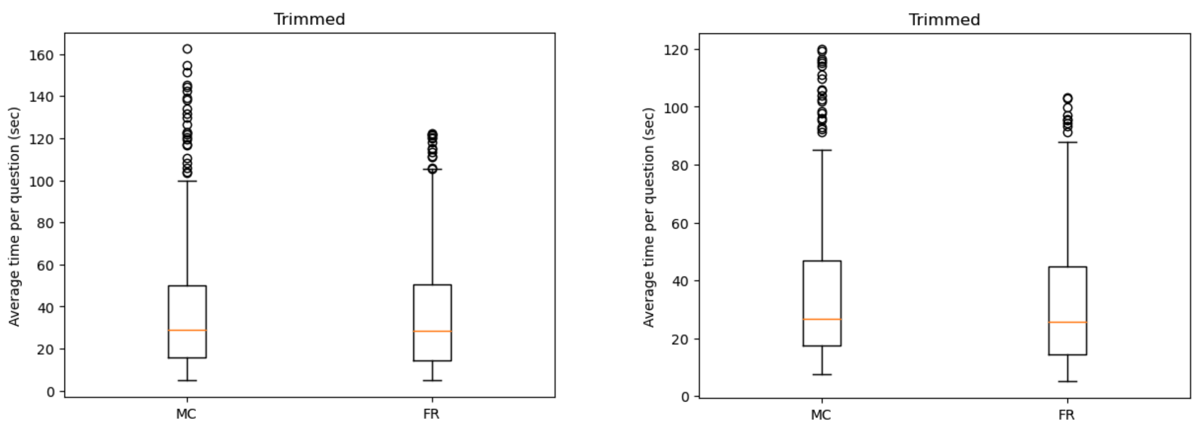
We measured the item response times, the time elapsed between the question appearing on the screen and the student submitting their answer. We trimmed the top and bottom 5% of response times, as the distributions are very positively skewed and there are many outliers. From Fig. 2, we can see that there is not a very large difference in response time between formats.

Mean Scores by Item Format

Term	Version	Pre/Post	n	Multiple-Choice		Free Response		t	d.f.
				$\bar{x}$	σ	$\bar{x}$	σ		
SP22	Algebra	Pre	451	53.4	25.8	25.9	32.8	14.0*	900
SP22	Algebra	Post	313	80.9	19.1	67.6	26.8	7.1*	624
SP22	Calculus	Pre	460	61.3	26.6	36.8	36.2	11.7*	918
SP22	Calculus	Post	319	85.5	15.0	73.7	22.3	7.9*	636
AU22	Algebra	Pre	715	43.4	27.8	24.1	30.6	12.5*	1428
AU22	Calculus	Pre	926	54.8	29.5	37.5	34.7	11.6*	1850

*$p < 0.001$

Table 1: Comparison of difficulty of multiple-choice and free response item formats.

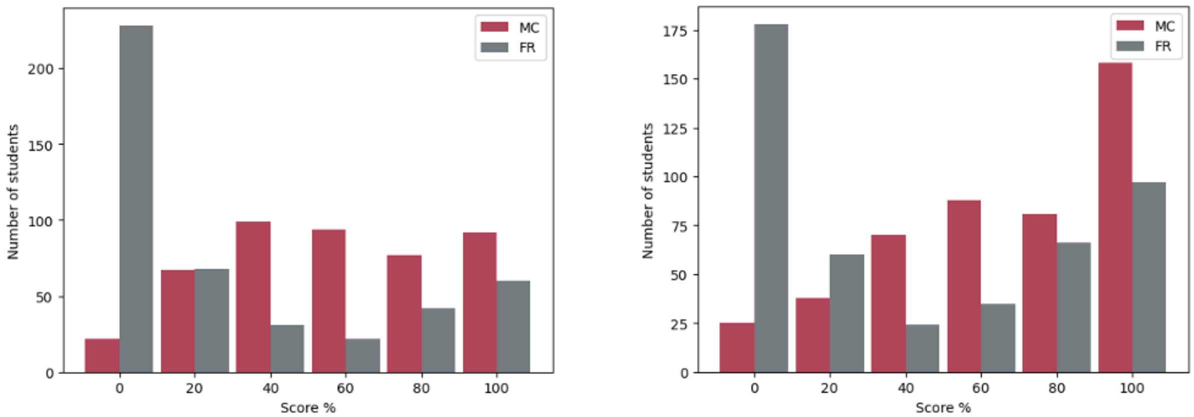


(a) Algebra-Based (b) Calculus-Based

Figure 2: Comparing students' average response time per item throughout the pre-test between answer formats in the SP22 term.

3.1.3 Discriminatory Power

To investigate whether there is a difference in discriminatory power between assessments in the two formats, we first examine the score distributions in each format on the pre-test. As shown in Fig. 3, the score distributions in both the multiple-choice and free response formats seem to be bimodal, for both the algebra-based and calculus-based students. In the free response format (shown in gray), these distributions have peaks at 0% and 100% and show clear distinctions between students who have mastered vector addition and those that have not. The distribution of scores in the multiple-choice format (shown in red) is harder to distinguish. We do see a peak at 100%, indicating that these students have mastered vector addition in the multiple-choice format, but the second peak is somewhere in the middle, and the two populations aren't as distinct. This indicates that assessments in the free response format have higher discriminatory power than assessments in the multiple-choice format.



(a) Algebra-Based

(b) Calculus-Based

Figure 3: Score distributions on the pre-test in the SP22 term.

To test this further, we calculated the upper-lower item discrimination index for each question on the pre and post-tests. Upper and lower groups are formed from the top and bottom 27% performing students. The item discrimination index is calculated by subtracting the number of students who got the item correct in the lower group from the number of students who got the item correct in the upper group, then dividing by the number of students in each group. From Table 2, we can see that in all cases, the free response format has a higher average item discrimination index than the multiple-choice format.

Average Item Discrimination by Item Format

Term	Version	Pre/Post	Multiple-Choice	Free Response
SP22	Algebra	Pre	0.67	0.84
SP22	Algebra	Post	0.32	0.63
SP22	Calculus	Pre	0.70	0.92
SP22	Calculus	Post	0.20	0.47
AU22	Algebra	Pre	0.58	0.63
AU22	Calculus	Pre	0.61	0.73

Table 2: Comparison of average upper-lower item discrimination index of multiple-choice and free response item formats.

3.1.4 Test Reliability

The next thing we would like to investigate is whether assessments in one format have higher reliability than assessments in the other format. We can measure this reliability with the reliability coefficient Cronbach's α. A criterion of 0.7 is traditionally used to determine whether an assessment has good reliability.

Table 3 shows that the free response format has higher reliability than the multiple-choice format in all cases. Additionally, the reliability is higher for the pre-tests than it is for the post-tests. This could be due to ceiling effects, where many students score very well on the post-test.

Test Reliabilty by Item Format

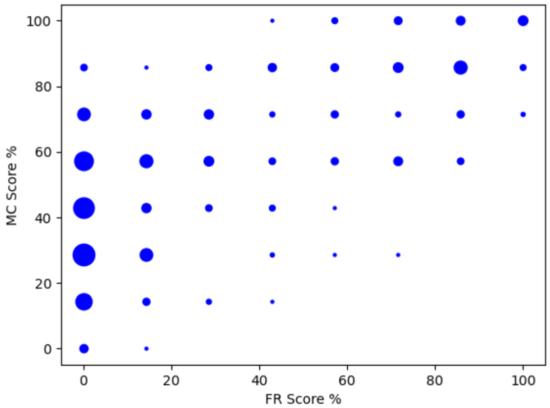
Term	Version	Pre/Post	Multiple-Choice	Free Response
SP22	Algebra	Pre	0.58	0.88
SP22	Algebra	Post	0.61	0.71
SP22	Calculus	Pre	0.66	0.88
SP22	Calculus	Post	0.49	0.59
AU22	Algebra	Pre	0.71	0.87
AU22	Calculus	Pre	0.76	0.87

Table 3: Comparison of Cronbach's α for multiple-choice and free response item formats.

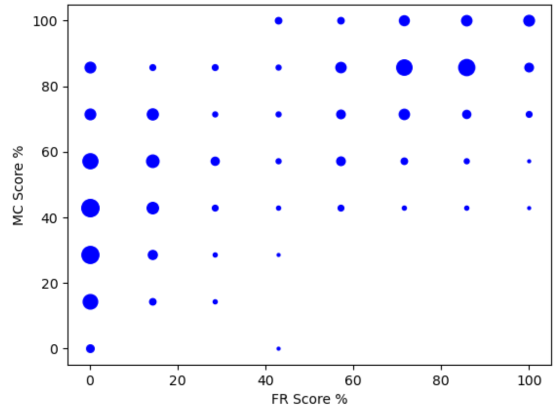
3.1.5 Hierarchy of Performance

A hierarchy of performance between the multiple-choice and free response formats would imply that answering correctly in one format would strongly predict answering correctly in the other, but not vice versa [2]. Figure 4 shows scatter plots of scores on the multiple-choice items vs scores on the free-response items. If a hierarchy of performance exists, we would see many more students in the top-left corner of the plot than in the bottom-right corner or vice versa. We do see this pattern where many students score well on the multiple-choice items and poorly on the free response items. However, very few students perform better on the free response items than on the multiple-choice items.

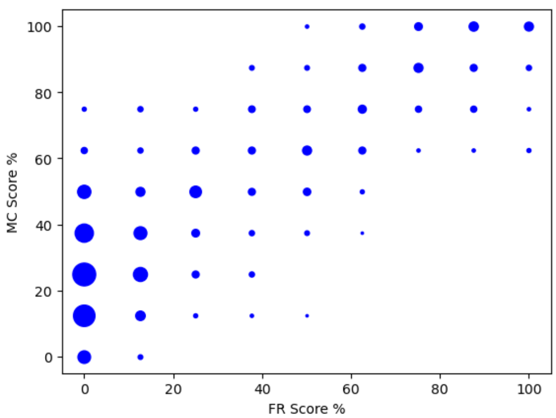
Note that this does not necessarily mean that understanding the multiple-choice format is necessary for understanding the free response format. Rather it is likely that as students are learning the concepts of vector addition, they can rule out certain incorrect answers to deduce the correct answer without explicitly knowing how to solve the problem. However, to correctly answer in the free response format, it is necessary to understand the process of answering vector addition questions completely.



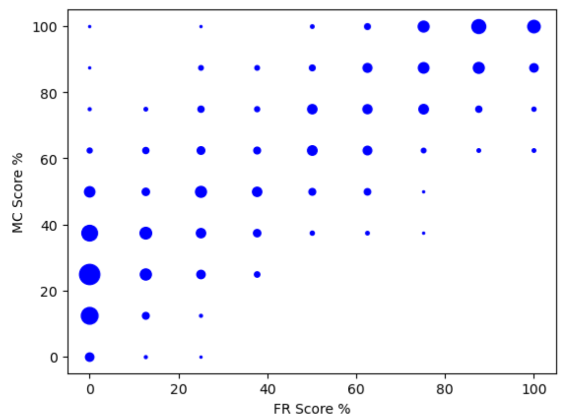
(a) SP22 Algebra-Based



(b) SP22 Calculus-Based



(c) AU22 Algebra-Based



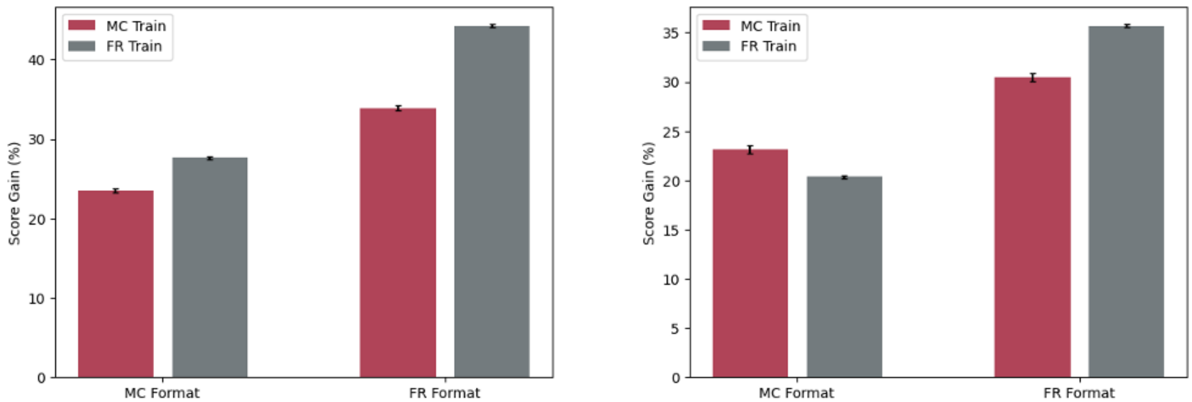
(d) AU22 Calculus-Based

Figure 4: Plots of multiple-choice scores vs free response scores. The area of each point is proportional to the number of students with that score combination.

3.2 Learning

3.2.1 Score Gains

Higher score gains (score on pre-test subtracted from score on post-test) in one answer format could indicate that training in that format is a better method for learning vector addition. Figure 5 compares score gains in both formats and between conditions. In both the algebra-based and calculus-based versions of the course, the FR training condition had higher gains in the items in the free-response format. In the calculus-based version, the MC training condition had higher gains in the multiple-choice format. Interestingly, in the algebra-based version of the course, the FR training condition had higher gains in the multiple-choice format than the MC training condition. Training in the free response format certainly seems to be better at preparing students to answer questions in the free-response format, but training in the free response format could also be better than training in the multiple-choice format at preparing students to answer questions in the multiple-choice format. This is investigated further in Section 3.2.3.



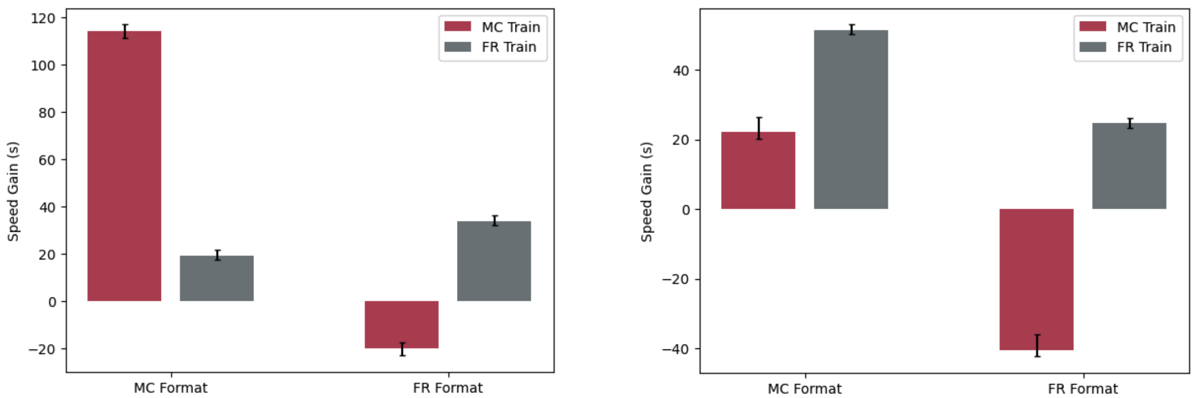
(a) Algebra-Based (b) Calculus-Based

Figure 5: Comparing score gains between training conditions in the SP22 term.

3.2.2 Speed Gains

When learning essential skills, we want the students to build fluency, which includes speed as well as accuracy. Figure 6 shows speed gains (time taken on post-test subtracted from time taken on pre-test) between formats and training conditions. Both training conditions sped up in the multiple-choice format in both versions of the course. For the multiple-choice items in the algebra-based version of the course, the MC training condition sped up much more than the FR training condition, but in the calculus-based version of the course, the FR training condition sped up more than the MC training condition. For the free response items in both versions of the course, the FR training condition sped up while the MC training condition slowed down.

It is unclear why the MC training condition slowed down on the free response items. One possible explanation is that these students had only used the format on the pre-test and needed extra time to figure out how to use the answer format. Another explanation is that now that most of the students have mastered vector addition, they have to take more time to be careful in selecting their answer whereas before they were less careful in answering. We saw that the most common incorrect answers in the free response format are some sort of averaging of the two initial vectors. This imprecise answering could support the latter hypothesis. Of course, there could be other explanations for this behavior, further investigation is needed to determine the cause.



(a) Algebra-Based (b) Calculus-Based

Figure 6: Comparing speed gains between training conditions in the SP22 term.

3.2.3 Modeling Effect of Training Condition

We created a multi-level model to predict a student's post-test score given their pre-test score, which training condition the student was in, and the format of the assessment. Eq. 1 shows the model we used. Pre_{ij} and $Post_{ij}$ are student i's pre-test score and post-test score, respectively, on the format j items. $Condition_i$ is set to 1 if student i was in the MC training condition and 0 if the student was in the FR training condition. $Format_j$ is set to 1 if the assessment is in the multiple-choice format and 0 if the assessment is in the free response format.

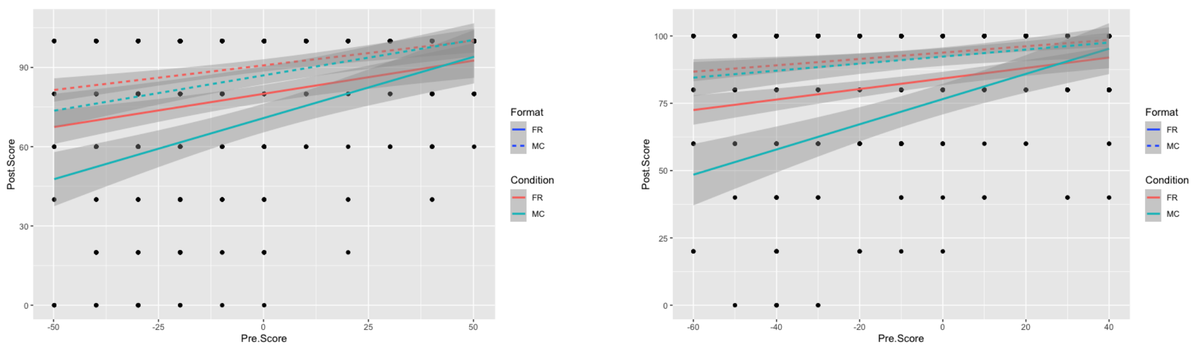
$$
\begin{aligned}
Post_{ij} = \beta_0 + \beta_{pre} * Pre_{ij} + \beta_{cond} * Condition_i + \beta_{form} * Format_j \\
+ \beta_{int} * Format_j * Condition_i + u_{0i}
\end{aligned}
\tag{1}
$$

Fitted Model Parameters

	β_0	β_{pre}	β_{cond}	β_{form}	β_{int}
Algebra-Based	80.02	0.29	-9.09	10.80	5.26
Calculus-Based	84.19	0.20	-8.54	9.53	7.37

Table 4: Fitted values of parameters in model (Eq. 1). Condition and format are set to 1 for MC and 0 for FR. Pretest scores are mean-centered.

The models for the algebra-based and calculus-based versions of the course were very similar. There was a significant positive β_{pre} term, which means that if the student gets a high score on the pre-test, they are more likely to get a high score on the post-test, which isn't surprising. There was a significant positive β_{form} term as well, which means that the student is more likely to get a high score if the assessment is in the multiple-choice format. This is also unsurprising, as we know that the free response format is more difficult. There also was a negative β_{cond} term, which means that the student is more likely to score highly if they were in the FR training condition. The interaction term β_{int} was also significant but was less significant than the rest of the terms.



(a) Algebra-Based

(b) Calculus-Based

Figure 7: Modeling post-test score vs pre-test score for each answer format and training condition.

From Fig. 7, we can see what this interaction term is doing. If the assessment is in the multiple-choice format, the negative β_{cond} term is partially canceled out by the positive β_{int} term for the MC training condition, meaning that the training condition matters very little. However, if the assessment is in the free response format, the β_{int} term is always 0, and the condition matters much more.

Also note that if a student scores highly on the pre-test, neither the format of the assessment nor the training condition matters, the student is very likely to score highly on the post-test regardless of the assessment format or their training condition. When the student does not score highly on the pre-test, then the training condition does matter if the assessment is in free response format. In assessments in the free response format for these students, those that practiced in free-response format will score higher than those that practiced in the multiple-choice format.

4 Conclusion

We see strong evidence that there are benefits to both learning and assessment of vector addition using the free response format in favor of the multiple-choice format. For assessing students' knowledge of vector addition, the free response format is more difficult, more discriminating, and more reliable than the multiple-choice format. Additionally, the free response format is hierarchically above the multiple-choice format, meaning correctly answering in the free response format strongly predicts being able to correctly answer in the multiple-choice format but the converse is not true.

The main goal we have for our students in practicing essential skills is to build fluency, including both speed and accuracy. For improving accuracy, we see that training in the free response format leads to higher score gains than training in the multiple-choice format. Training in the free response format leads to similar score gains in the multiple-choice format than training in the multiple-choice format, but training in the multiple-choice format does not lead to similar gains in the free response format to training in the free response format. We also see that training in the free response format leads to speed-ups in both formats. Training in the multiple choice format also leads to speed-ups in the multiple-choice format but concernedly leads to slowing down in the free response format.

For many assessments, instructors will choose the multiple-choice format for ease of grading. On paper, it takes significantly longer to grade free response items. However, in the online format, both formats are graded automatically, and the free response items are just as easy or easier to create than the multiple-choice items. It also doesn't significantly slow down the students to use the free response format, the response times are similar between formats. These usual trade-offs between multiple-choice and free response answer formats are of no concern for online vector addition practice. Thus, for online-based training and assessment of vector addition, the free response format seems to be better than the multiple-choice format.

References

[1] R. D. Knight, "The vector knowledge of beginning physics students", The physics teacher **33**, 74–77 (1995).

[2] A. F. Heckler and T. M. Scaife, "Adding and subtracting vectors: the problem with the arrow representation", Physical Review Special Topics-Physics Education Research **11**, 010101 (2015).

[3] N.-L. Nguyen and D. E. Meltzer, "Initial understanding of vector concepts among students in introductory physics courses", American journal of physics **71**, 630–638 (2003).

[4] U. Wutchana and N. Emarat, "Students' understanding of graphical vector addition in one and two dimensions", International Journal of Physics & Chemistry Education **3**, 102–111 (2011).

[5] S. Flores, S. E. Kanim, and C. H. Kautz, "Student use of vectors in introductory mechanics", American Journal of Physics **72**, 460–468 (2004).

[6] P. S. Shaffer and L. C. McDermott, "A research-based approach to improving student understanding of the vector nature of kinematical concepts", American journal of physics **73**, 921–931 (2005).

[7] S. F. García, C. Cuellar, M. D. G. Quezada, O. R. Sandoval, M. d. l. Á. C. Quiñones, and V. Aguirre, "Problemas de entendimiento conceptual con las operaciones entre vectores", Latin-American Journal of Physics Education **11**, 7 (2017).

[8] B. Latifa, E. Purwaningsih, and S. Sutopo, "Identification of students' difficulties in understanding of vector concepts using test of understanding of vector", in Journal of physics: conference series, Vol. 2098, 1 (IOP Publishing, 2021), p. 012018.

[9] P. Barniol and G. Zavala, "Vector addition: effect of the context and position of the vectors", in Aip conference proceedings, Vol. 1289, 1 (American Institute of Physics, 2010), pp. 73–76.

[10] J. M. Hawkins, J. R. Thompson, M. C. Wittmann, E. C. Sayre, and B. W. Frank, "Students' responses to different representations of a vector addition question", in Aip conference proceedings, Vol. 1289, 1 (American Institute of Physics, 2010), pp. 165–168.

[11] S.-Y. Lin and C. Singh, "Can free-response questions be approximated by multiple-choice equivalents?", American Journal of Physics **81**, 624–629 (2013).

[12] A. D. Slepkov and R. C. Shiell, "Comparison of integrated testlet and constructed-response question formats", Physical Review Special Topics-Physics Education Research **10**, 020120 (2014).

[13]M. E. Martinez, "A comparison of multiple-choice and constructed figural response items", Journal of Educational Measurement **28**, 131–145 (1991).

[14]S. L. Wise and X. Kong, "Response time effort: a new measure of examinee motivation in computer-based tests", Applied Measurement in Education **18**, 163–183 (2005).