Alterations in Student Performance with In-Person, Remote Synchronous, and Hybrid Learning During COVID-19 Pandemic

J. Weinberger, A. Opalka, T. May, and N. Chernets

Drexel University School of Education and Drexel University Graduate School of Biomedical Sciences and Professional Studies

Abstract

The COVID-19 pandemic has altered the instructional modalities utilized within various graduate school levels. The quick implementation of in-person, remote synchronous, and hybrid lecture styles has created a need to assess how student performance compares with these various instructional modalities. To understand how the widespread shift to remote instructional modalities has affected student learning outcomes, we compared student outcomes for the same exam across 2018-2021. We chose the first exam of a first-year Molecular Structure and Function course (Core 1) taken by all first-year biomedical sciences M.S. and Ph.D. students and assessed alterations of student success and shortcomings. We correlated differences in student performance with the different instructional modalities, in-person, remote synchronous, and hybrid. Additionally, we evaluated multiple choice questions of the exam and made recommendations to improve exam quality. This completed assessment provides adequate knowledge of how the COVID-19 pandemic has been a catalyst for a transition to remote synchronous instruction and learning and measures its success.

Research Description

- Research Purpose
  - To evaluate student performance of first-year biomedical sciences M.S. and Ph.D. students on Exam 1 of Molecular Structure & Function (Core 1) with varied instructional modalities before and during the COVID-19 pandemic.
- Problem Statement
  - The COVID-19 pandemic necessitated a widespread shift to remote instructional modalities; however, little is known regarding how this shift affected student learning outcomes. This research is designed to fill this gap and better understand how student performance changed at various stages of the pandemic due to different instructional modalities (in-person, remote synchronous, and hybrid).
- Research Questions
  - Did a switch to remote synchronous learning cause changes to student performance on Exam 1 of a Molecular Structure and Function course taken by first-year biomedical science M.S. and Ph.D. students?
  - Did performance on the exam correlate with the instructional modality used?
- Findings
  - Following the changes in instructional modalities from in-person to remote synchronous to hybrid, student learning outcomes remained consistent. Alterations in the level of difficulty remained the same, and no additional outliers were created due to the change in teaching format.

Methods

212 students that completed Exam 1 of Core 1 between 2018 and 2021 were included in this assessment. Students that were enrolled as first-year M.S. and Ph.D. students in biomedical sciences were required to take the Molecular Structure and Function course. This course was team-taught by researchers who provided 10 points of questions, either multiple-choice or open-ended. This assessment focused on multiple-choice questions that remained identical across the years, which resulted in 25 multiple-choice questions.

Student Performance Evaluations: Exam 1, Core 1 (*N*=212)

Cohort Years	Number of Students
2018	58
2019	45
2020	49
2021	60

Note. 25 multiple-choice questions were used for comparisons; both questions and answer choices remained unchanged from year to year. Question order changed between 2019 and 2020.

Results

Item Difficulty

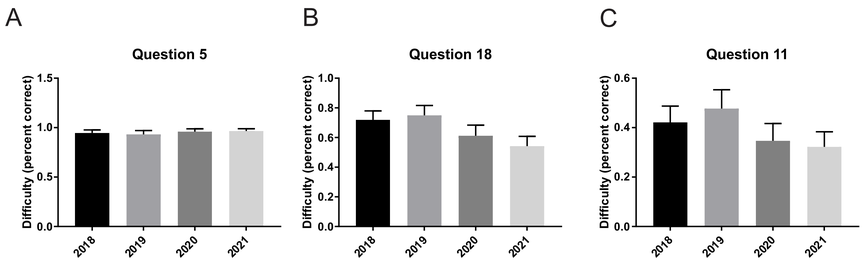

Figure 1: A. Low item difficulty question (p-value >0.8). **B.** Moderate item difficulty question (p-value 0.5-0.8). **C.** High item difficulty question (p-value <0.5).

To determine changes across student performance between years 2018 through 2021, item difficulty was calculated as the percentage of correct answers out of the total number of student answers. Low item difficulty was defined as >80% correct (p-value >0.8), which consisted of ten out of 25 multiple-choice questions. Moderate item difficulty was defined as 50–80% correct (p-value 0.5–0.8), which consisted of seven out of 25 multiple-choice questions. High item difficulty was defined as <50% correct (p-value <0.5), which consisted of eight out of 25 multiple-choice questions. Therefore, questions with higher p-values were considered easier, while questions with lower p-values were considered harder. A one-way ANOVA revealed no significant differences between years, which illustrates that the difficulty is consistent across years, irrespective to instructional modalities.

Item Discrimination

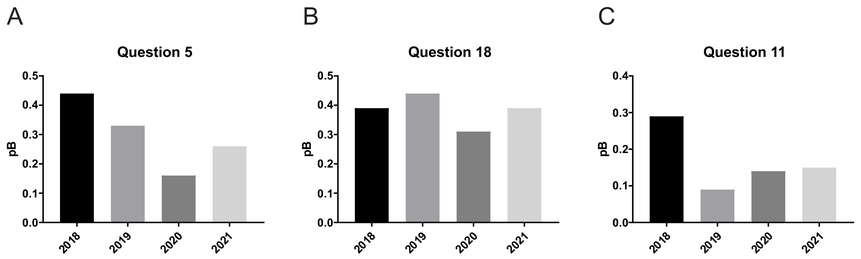

Figure 2: A. Low difficulty question (p-value >0.8). **B.** Moderate difficulty question (p-value 0.5-0.8). **C.** High difficulty question (p-value <0.5).

To investigate item discrimination between scores on a single item and the total test score per student, point-biserial (pB) correlation coefficients were calculated. The relationship between these items and the overall test allowed for correlations between students who scored highly on the test and students who are most likely to answer the question correctly. Questions with positive pB values are a sign of proper item discrimination between high-scoring and low-scoring students, while questions with negative pB values are considered to show poor item discrimination between high and low scoring students. **Figure 2A** shows a low item difficulty question (p-value >0.8), **Figure 2B** shows a moderate item difficulty question (p-value 0.5–0.8), and **Figure 2C** shows a high item difficulty question (p-value <0.5), where all cohort years show positive pB values. The lack of significance between the cohort years when compared via a one-way ANOVA illustrates that there are no differences between instructional modalities. Positive pB values further substantiate the proper item discrimination between the high-scoring and low-scoring students.

Item Distractor

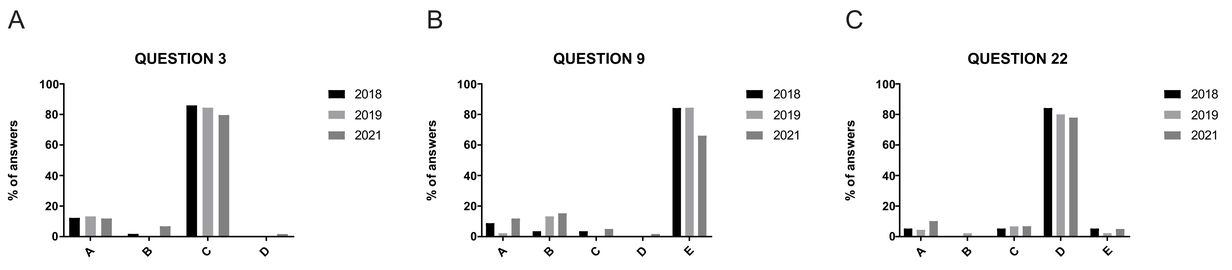

Figure 3: A. Low item difficulty question; correct answer choice ‘C’. **B.** Moderate item difficulty question; correct answer choice ‘E’. **C.** High item difficulty question; correct answer choice ‘D’. Individual answer choices were not available for 2020 cohort.

To investigate the correlation between cohort years and student answer choices, an item distractor analysis was performed. The percentage or occurrence for each answer choice is illustrated to identify potential item distractors for students. Item distractors with high occurrence can correlate with an increased likelihood of student guessing. Questions with low occurrence of item distractors indicate a more consistent degree of student performance. **Figure 3A** shows a low item difficulty question, **Figure 3B** shows a moderate item difficulty question, and **Figure 3C** shows a high item difficulty question. The lack of significance between the cohort years when compared via a one-way ANOVA illustrates that there are no differences between instructional modalities. A lack of high yield item distractors further substantiates no differences between instructional modalities and student answer choices.

Discussion

Following the required changes in instructional modalities from in-person (2018, 2019) to remote synchronous (2020) to hybrid (2021) during the COVID-19 pandemic, performance by M.S. and Ph.D. biomedical sciences students on their first exam in Molecular Structure and Function remained consistent. This was indicated by no significant differences in item difficulty, discrimination, or distractors on 25 identical multiple-choice questions, irrespective of instructional modalities.

Item difficulty analysis revealed no significant differences across years, and the set criteria divided the questions almost evenly into low, moderate, and high item difficulty. Although there were no significant differences in item difficulty across each cohort year, it is important to highlight some findings amongst item difficulty. First, more than a third of questions were considered low item difficulty, yet this could be due to a bias in our sample that consisted of M.S. and Ph.D. biomedical sciences students. For example, these students may have a higher general aptitude or motivation to grasp moderately difficult concepts rather than the exam testing easier content. Second, most high item difficulty questions could be distracting test takers rather than testing the content due to negatives in the question (two out of eight questions), all-or-none answer choices (two out of eight questions), and variation in length of answer choice (three out of eight questions). Therefore, to ensure the questions are accurately testing the content, revising questions and answer choices may benefit student performance.

Item discrimination analysis revealed no significant differences across years and resulted in a pB score >0.2 for sixteen out of 25 questions (and >0.1 for 22 out of 25 questions). Only two questions received lower than <0.1 and one questioned received <0.0. Of these three questions, two were indicated as low item difficulty and one as high item difficulty. Nonetheless, most questions across cohorts had positive pB values, which demonstrates proper item discrimination between the high-scoring and low-scoring students across the cohort years and instructional modalities.

Item distractor analysis revealed no significant differences across cohort years, which indicates no differences between instructional modalities. However, this analysis revealed that students were between two answers on seven out of 25 questions and three answers on 4 out of 25 questions (11 out of 25 questions total). This could indicate that students may be guessing on certain questions, whether due to unclear question/answer choices or unclear content.

Recommendations

Collectively, items are discriminating well across years, irrespective of learning environment. It is important to note that this course is taken by students at two different campuses with the option of video conferencing half the lectures or traveling to the other campus to listen to the lecturer on the site (years 2018, 2019, and 2021). Most students choose to remain at their campus regardless of where the lecturer was speaking. Due to this hybrid model of instruction already existing, this indicates that learning environment, if the lecture is live, may not influence student performance.

However, it is important to note that the average score on multiple choice questions across three years (not available for 2020) was 72.08%. Since 80% correct is required for some of these students’ biomedical sciences programs, it may be beneficial to revise the wording to improve this average and/or item discrimination between high- and low-scoring students.

Acknowledgements

This project was supported by Burroughs Wellcome Fund Career Guidance for Trainees Grant #CGT2110275.