EDUCATIONAL RESEARCH

HAPS EDUCATOR
Journal of the Human Anatomy and Physiology Society

Volume 28, Issue 2 | August, 2024
https://doi.org/10.21692/haps.2024.008

Answer Changing Behaviors and Performance in a First-Year Medical Gross and Developmental Anatomy Course

Marli Crabtree, MS[1], Kenneth L. Thompson, PhD[1], Ellen M. Robertson, PhD, OTR/L[2]

[1]Department of Advanced Biomedical Education, University of Mississippi Medical Center, Jackson, MS, USA

[2] Department of Physician Assistant Studies, Randolph-Macon College, Ashland, VA, USA

Corresponding author: ellenrobertson@rmc.edu

Abstract

Research has suggested that changing one's answer on multiple-choice examinations is more likely to lead to positive academic outcomes. This study aimed to further understand the relationship between changing answer selections and item attributes, student performance, and time within a population of 158 first-year medical students enrolled in a gross and developmental anatomy course at an academic medical institution in the United States. For each student, answer changes, overall exam performance, and individual item performance data were retrieved from an online testing software for a single block exam. Researchers determined how many times students changed their answers, the associated outcomes, and time spent on each item and the entire exam in relation to item performance. Students in the highest performing quartiles were more likely to keep their initial answer selection, spent more time choosing their initial answer selection, and averaged a higher total exam time than each of the lowest two performance quartiles. Time on individual items and answer changes had a statistically significant relationship, with more time relating to the presence of an answer change. Changing an answer selection was more likely to result in a negative outcome. The content subject was significant in relation to answer changes and time spent per item. This study provides a deeper understanding into which factors, such as item attributes, time, and performance of the student, showed statistically significant relationships to answer changing. https://doi.org/10.21692/haps.2024.008

Key words: changing answers, test-taking practices, developmental anatomy, gross anatomy, medical education

Introduction

Answer Changing Research

As early as 1929, Mattews found that 86% of college students believed that changing their answer led to an overall negative outcome, and, in 1984, Benjamin et al. similarly found that 55.2% of faculty believed that changing their answer would lead students to a negative outcome. This belief has been seen across decades and is still present in classrooms today (Benjamin et al., 1984; Cox-Davenport et al., 2014; Merry et al., 2021).

Much of the existing literature observes nursing, psychology, and undergraduate student populations, with fewer studies observing medical and dental students. Ultimately, the studies that did observe these professional degree students came to similar conclusions. In 2017, Pagni et al. found that 99.4% of dental students in their study benefitted from answer changing. Ferguson et al. (2002) studied second-year medical students and the impact of answer changes on their performance in a foundational science course and concluded that changing one's answer from the initial selection resulted in a significant positive change. Answer-changing research in the medical student population has varied in their aims, with some addressing the impact of personal preferences like learning styles on answer changing and others strictly observing the outcomes associated with changing an answer (McNulty et al., 2007; Merry et al., 2021). Similar to other student populations, this research has shown that changing answer selections is most commonly beneficial, but because of the various study designs and aims found in existing literature, more research is needed to better understand the answer changing behaviors of medical students.

Factors Related to Answer Changing

Although research on answer changing behaviors exists, and the outcomes regarding answer changing are fairly consistent, there is less understanding and agreement regarding the

continued on next page

factors that relate to the presence and pattern of answer changes. Over time, researchers have hypothesized that factors like student proficiency in the associated courses and/or on the associated exams, item attributes, the order of questions and time allotment, and other possible variables could be influencing students to exhibit the answer changing behaviors they do (Ferguson et al., 2002; Geiger, 1997; Harvill & Davis, 1997; Liu et al., 2015; Mueller & Wasser, 1977; Ouyang et al., 2019; Pettijohn II & Sacco, 2007).

Earlier studies that addressed the impact of student proficiency on answer changes had varying outcomes, with some showing statistical significance and others showing insignificance between the behaviors of the highest and lowest performing students (Ferguson et al., 2002; Friedman & Cook, 1995; Mueller & Wasser, 1977). A study completed in 2021 concluded there were no indications that students of different proficiency levels were more likely to benefit from or be harmed by answer changing, and multiple older studies provided the same conclusions (Archer & Pippert, 1962; Matthews, 1929; Merry et al., 2021). Contrastingly, a study completed by Ouyang et al. (2019) considered student proficiency and found the opposite. This study divided students into higher, medium, and lower academic ability groups based on their performance on a high-stakes exam. Students who were determined to be of higher academic ability were found to review more items, were less likely to change answers, and were more likely to make an incorrect to correct change than students of lower academic ability (Ouyang et al., 2019). Other studies have also found that answer changes were more beneficial to higher performing students (Ferguson et al., 2002; Harvill & Davis, 1997). Item attributes, such as the presence of a picture, subject content, and recency of the content used (material in the current block or review material from previous blocks) have been less commonly studied in relation to answer changing.

The relation between time and answer changing was considered by Ouyang et al. (2019). The researchers found that time taken on an item had significant relationships with pattern of change and varied between students with different proficiency levels (Ouyang et al., 2019). This study found that students spent the most time on correct to correct changes (presumably, changes from correct to incorrect and then another change from incorrect to correct) followed by incorrect to incorrect changes (Ouyang et al., 2019). Further, students in this study spent the least amount of time on correct to incorrect and incorrect to correct changes (Ouyang et al., 2019). In 2002, Ferguson et al. found that higher item times were significantly related to the presence of an answer change, with more time resulting in a higher likelihood of an answer change. The authors found no other studies that researched the relationships between time spent on each individual item in relation to the presence and pattern of answer changes. Likewise, few studies expanded beyond correct to incorrect, incorrect to correct, and incorrect to incorrect patterns of change students made on individual items (Ouyang et al., 2019).

Goals of this Study

This study aimed to fill a gap in the existing literature on answer changing behaviors. Looking at first-year medical students in a gross and developmental anatomy course, this study aimed to observe the presence, patterns, and outcomes of answer changes in relation to specific item attributes on a multiple-choice exam. This study dove deeper into patterns of answer changes, looking at not only single changes but also multiple answer changes. Another aim of this study was to determine if time spent on individual items and the exam as a whole was related to the presence and pattern of changes made by students. Comparatively, the researchers also aimed to identify differences in the answering behaviors of high and low performing students.

Materials and Methods

This study was approved by the Institutional Review Board (IRB) at the University of Mississippi Medical Center, UMMC-IRB-2022-356.

Educational context

This retrospective and descriptive study was completed using data from an academic medical center in the southeastern United States. The gross anatomy (GA) and developmental anatomy (DA) course was region-based and organized into content blocks. Block one covered the back and upper limb, block two covered the thorax and abdomen, block three covered the pelvis and lower limb, and block four covered the head and neck. Anonymous performance data was obtained and studied from the third block summative multiple-choice exam for 158 students enrolled in their first year. The block three exam was the only assessment observed in this study. Because of software and data limitations, only one exam could be observed. The third block exam was chosen due to its number of review and image items, as well as the broad array of anatomical content covered in the pelvis and lower limb block (musculature, vasculature, neuroanatomy, organs, soft tissues, and bony elements).

The third block exam was administered synchronously on campus using the exam-taking software ExamSoft (https://examsoft.com/). The exam was proctored and completed on students' personal devices within a locked browser. ExamSoft logged the activity of each student who took this exam electronically, creating data that provided time stamps, movement between questions, answer changes, and outcomes for each item. Students were allotted a total of three hours or 180 minutes to complete the exam. Although there was no access to student accommodation information to determine whether any students were approved for extra exam time, all students in this study completed the exam within the standard 180-minute time limit. Consequently, potential accommodations were not considered as criteria for exclusion. Students who completed the exam on paper were excluded from the study, as were three graduate students enrolled in the course.

continued on next page

This block exam consisted of 100 questions. Three questions were excluded due to the items being omitted before final grades were calculated, leaving a total of 97 questions used in data analysis. Of the 97 items, there were 65 GA and 32 DA questions. The items were a mix of first and second order items, with some being in the form of clinical vignettes. Nineteen questions incorporated images. The image-based questions consisted of radiology imaging, cross-sectional anatomy, DA depictions, and questions based on diagrams or clinical imagery. The exam also incorporated review content from previous blocks, consisting of 16 total review items, nine covering GA content and seven covering DA content. All review items were newly crafted questions, and none were exact copies of questions students had seen on either of their two previous exams. The course director and instructors completed an extensive review of the exam items and removed all negatively coded and worded items before the exam was distributed to students. All items in this exam had four answer options, and none of the options were "all of the above" or "none of the above."

Data collection

For each student, an exam activity report was downloaded from ExamSoft that contained time stamps, navigation between questions, answer changes, and outcomes for each item. Activity data were then de-identified by removing all student identification numbers and assigning each student a random number from five to 163. The data for each student were then combined and stored within IBM SPSS (Statistical Package for the Social Sciences) Statistics for Windows (2020, Version 28), where all data analysis was later completed. The data collected and observed from ExamSoft included the individual answering behaviors of each student.

Answering behavior

The specific student answering behaviors observed were the correctness of their initial answer selection, presence of a change or no change to their initial answer, the pattern of change, if any, and associated outcome, either correct or incorrect. In addition, the amount of time the student spent per item and the total time spent on the exam were determined. The determination of the time spent on each item, time spent on the total exam, and answer changes made were completed using the exam activity reports produced by ExamSoft.

The final response for each item was categorized as not changed (the student kept their initial answer selection) or changed. Additionally, the pattern of change for each item was determined as one of the following: no change in answer (0), change from the correct answer to an incorrect answer (1), change from an incorrect answer to the correct answer (2), change from an incorrect answer to a different incorrect answer (3), multiple changes (initial answer choice was correct or incorrect) ending with the correct answer (4), or multiple changes (initial answer choice was correct or incorrect) ending with an incorrect answer (5).

The correctness of each item was determined using an answer key and the activity report for each student. For each item, a timestamp was recorded documenting the time at which the student navigated away from the item. To determine the time on an item, the time at which the student navigated away from the previously visited item was subtracted from the time at which the student navigated away from the current item. If a student revisited an item, the time spent on each revisit was added to the total item time. Total exam time for each student was calculated by subtracting their start time from their submission time. Total exam times were rounded to the nearest minute. Item times were reported in either seconds only or minutes and seconds (minutes.seconds), which will be explained where relevant.

Item Attributes

Data regarding item attributes included the recency of the content (if the item was current block material or review content from a previous block), the subject content of the item (DA or GA), and whether the item utilized a picture or only consisted of words. To determine item attributes, a blank copy of the exam was observed and attributes for each item were recorded. This categorization was completed by the first author, and once completed, reviewed by the last author to confirm accuracy.

Data Analyses

Statistical analysis was completed using multiple statistical methods with the alpha value set to 0.05 for all statistical tests used. For the purposes of this study, the mention of significance is referring to statistical significance. A mix of chi-square tests of independence, independent samples t-tests, and analysis of variance (ANOVA) tests was used for data analysis and will be discussed where the respective results are reported.

Results

Descriptives

Students were classified into quartiles based on their total exam performance on the third block exam. All exam scores reported are in the form of percentages, with a minimum of zero and a maximum of 100. Total exam scores ranged from 53-94, with quartile four (Q4) being the lowest performing students and quartile one (Q1) being the highest performing students. The average times spent per item and on the total exam were calculated for each quartile of students. Student quartiles were used to determine differences in answering behaviors. The range of exam scores used to determine the quartile each student was categorized into can be found in Table 1.

Quartile Differences

An ANOVA compared the time per item to each performance quartile, and this relationship was found to be significant ($F(3, 15{,}419) = 6.788$, $p < 0.001$). Item time averages and ranges for each performance quartile can be found in Table 1. Because the ANOVA assumption of homogeneity of variances was violated, a Games-Howell post-hoc correction was used to identify

continued on next page

significant differences between quartiles. Time spent per item was significant when comparing Q4 and Q2 ($p < 0.001$), with an average difference of eight seconds. Time spent per item was also significantly different by 5 seconds when comparing Q3 and Q2 ($p = 0.025$).

An ANOVA was also conducted to test the relationship between performance quartile and the total time spent on the exam. The ANOVA identified a significant difference ($F (3, 15,419) = 115.658$, $p < 0.001$). Exam time ranges and averages for each performance quartile can be found in Table 1. A Games-Howell post-hoc correction identified significant differences in average exam times between each quartile ($p < .001$).

To analyze the relationship between the performance quartile and both the presence and pattern of change, a chi-square test of independence was used. The relationship between performance quartile and the presence of an answer change was significant ($X^2 (3, N = 15,423) = 92.496$, $p < 0.001$). Quartile one logged the highest number of answers that had not been changed. Conversely, Q4 averaged the most answer changes. Quartiles two and three were similar in that both groups changed their answer more often than keeping their initial selection. Quartile one was the only quartile that kept their initial selection at a higher rate than changing their answer.

There was also a significant relationship between the performance quartile and the pattern of change made ($X^2 (15, N = 15,423) = 102.213$, $p < 0.001$). Quartile four, the lowest performing quartile, logged the most changes from correct to incorrect compared to all other quartiles. Furthermore, Q4 had the highest number of multiple answer changes ending with an incorrect selection, with nearly two times as many of these changes when compared to Q1. Multiple answer changes to the same item were most beneficial to Q1. Quartiles two and three were similar in logged answer-changing patterns, showcasing nearly identical numbers of incorrect to correct answer changes and multiple changes ending with an incorrect answer.

Time and Item Attributes

ANOVA indicated a significant difference between the time spent on an item and the pattern of change ($F (5, 15,417) = 552.001$, $p < 0.001$). The average item times associated with each pattern of change are presented in Table 2.

Quartile	Score ranges	Item time average (sec)	Item time ranges (min.sec)	Exam time avg (min)	Exam time ranges (min)
Q1	83-94	63	00.03-32.35	108	47-177
Q2	78-82	67	00.03-15.19	112	66-180
Q3	74-77	62	00.03-16.38	104	38-180
Q4	53-73	59	00.02-12.36	100	38-180

Table 1. *Quartile differences: score ranges and item and exam time averages and ranges.*

	Item time average (min.sec)	P-value
Pattern of change		< .001
No answer change	0.51	
Correct to incorrect	2.03	
Incorrect to correct	0.45	
Incorrect to incorrect	2.02	
Multiple changes, end correct	2.29	
Multiple changes, end incorrect	0.54	

Table 2. *Presence and pattern of change in relation to average item and total exam times*

continued on next page

Using an independent samples t-test, a significant relationship was found between time spent on each item and the correctness of the final answer selection (t (15,421) = 23.118, $p < 0.001$). The assumption of equal variances was violated based on the Levene's Test ($p < .001$). The average item times for these relationships are presented in Table 3.

An independent samples t-test analyzed the significance between the time spent on an item and item attributes, including the presence of a picture, subject content, and recency of the content. Significance was determined between item time and the presence of a picture (t (15,421) = 9.123, $p < 0.001$). The assumption of equal variances was violated based on the Levene's test ($p < 0.001$). Additionally, significance was determined between item time and subject content (t (15,421) = 33.870, $p < 0.001$). The Levene's test indicated that the assumption of equal variances was violated ($p < 0.001$). In contrast, item time was not significant in relation to the recency of the item content (t (15,421) = -.400, $p = .694$). All average item times in relation to item attributes are presented in Table 3.

	Item time average (min.sec)	P-value
Correctness		< .001
Correct	0.54	
Incorrect	1.34	
Presence of picture		< .001
Picture	1.14	
No picture	1.00	
Subject content		< .001
Developmental anatomy	0.34	
Gross anatomy	1.17	
Recency of content		.694
Review	1.03	
Current	1.03	

Table 3. *Item Outcomes and attributes: item time averages and significance*

Presence and patterns of change

A chi-square test of independence was conducted to test for significance between the presence of an answer change and the outcome of the selection and found significance (X^2 (1, N = 15,423) = 600.11, $p < 0.001$). Answer selections that were incorrect were more likely to have been changed compared to correct answer selections.

There was a significant relationship between the subject content of the item and the presence of an answer change, as determined by a chi-square test of independence (X^2 (1, N = 15,423) = 756.086, $p < 0.001$). There were nearly two times the amount of answer changes on GA items compared to DA items. On DA items, 63% of responses had not been changed. Furthermore, a significant relationship was identified between the subject content of the item and the pattern of change (X^2 (5, N = 15,423) = 118.324, $p < 0.001$). There were nearly twice as many correct to incorrect, multiple changes ending with a correct response, and multiple changes ending with an incorrect response on GA items compared to DA items.

Chi-square tests of independence found no statistical significance between the presence of a picture and the recency of the item content in relation to the presence or pattern of an answer change.

continued on next page

Discussion

Research conducted on the topic of changing one's answer has consistently shown that it is largely beneficial to do so (Archer & Pippert, 1962; Coffey et al., 2024; McMorris et al., 1987; Ouyang et al., 2019; Pagni et al., 2017). The results of this study contradict these findings.

Breaking down this medical student sample into performance quartiles allowed for a deeper understanding of the impact of answer changing on outcomes. The highest performing students changed their answer less overall than any other performance quartile. When students did make an answer change, it was most commonly beneficial to high performing students. This finding differs from much of the existing literature on the impact of student proficiency (Archer & Pippert, 1962; Matthews, 1929; Merry et al., 2021), but supports more recent findings in health professional student populations (Miller et al., 2023; Ouyang et al., 2019), that changing answers is a more beneficial practice for high performing students.

Although many studies have observed the impact of binary answer changes (correct to incorrect, incorrect to correct, incorrect to incorrect), few studies had further examined the impact of multiple answer changes to a single item (Coffey et al., 2024; Ouyang et al., 2019). Because this study sorted individual item answer changes into single and multiple patterns, it was possible to establish that Q1 students were the most common performance quartile to make a single change from incorrect to correct. Adversely, Q4 students were nearly two times as likely than all other performance quartiles to change their answer multiple times and end with an incorrect selection.

The number of times a change was made, single or multiple, and the outcome of those changes may provide valuable insight into the differences in answering behaviors between these student samples. These findings suggest that higher-performing students might be more confident in their decision-making process, resulting in an educated, single change, while lower performing students could struggle with higher levels of self-doubt invoked by a lack of knowledge or confidence. Students that doubt their knowledge or feel unconfident would likely be more prone to guess an answer and change their guess multiple times, as confidence is a theme that has emerged in previous answer changing research (Cox-Davenport et al., 2014; Stylianou-Georgiou & Papanastasiou, 2017). Encouraging students to 'revisit and rethink' their answers if they feel unsure might be beneficial to high performing students, but for those who are lower performing, it could be harmful. Literature has shown that advice from faculty does have an impact on students' answer changing practices during exams, further highlighting the importance of how educators are addressing and advising their students as a whole (Bauer et al., 2007; Cox-Davenport et al., 2014; Merry et al., 2021).

This study also considered the relationship between time spent on an item and the presence and pattern of change as well as the outcome of those changes. When analyzing all responses together, a statistically significant relationship was found between time spent on an item and the presence of an answer change, with more time significantly related to the presence of a change. Additionally, it was found that longer item times were more commonly associated with incorrect answer selections. This suggests that the longer students spend on an item, the more likely they will be to change their answer and get the answer incorrect, as has also been found in previous research (Ouyang et al., 2019).

Given the significant relationships for the entire study sample regarding time and outcomes, the findings were not as expected when studying the same relationship with individual performance quartiles. The data showed that Q1 and Q2 students averaged the highest exam and individual item time averages but were the least likely to change their answer. In contrast, Q4 students averaged the lowest exam times, lowest average time spent on individual items and were the most likely group to change their answer. These findings showcase how different the answer changing behaviors were for the various performance quartiles, and how studying the sample as a whole led to results that did not represent the practices of the highest and lowest performing student groups. The difference in these findings might suggest that the highest performing students are much more intentional in their decisions and have higher metacognitive ability (Stylianou-Georgiou & Papanastasiou, 2017). By averaging higher times and lower numbers of answer changes, their choices seem to be more confident when compared to the behaviors of the lowest performing students. This, again, leads to the assumption that encouragement to either change or not change answers would likely not be beneficial to all students in a classroom.

Multiple previous studies found that item format showed significance in relation to answer-changing practices, but few other studies had considered other item attributes, such as the inclusion of a picture and specific content covered (Fischer et al., 2005; Geiger, 1997; Harvill & Davis, 1997). The results of this study suggest that not only is there significance between item attributes and answer changing, but that there are differences in the impact each specific item attribute can have on the presence and pattern of answer changes.

When considering the impact of the item content on time, there was a significant difference in the time spent on GA items compared to DA items, with GA items taking 43 seconds longer, on average, to be answered. Furthermore, answer selections on GA items were changed nearly twice as often as DA items. Although the presence of change was higher on GA items, the data showed that the most frequent pattern of change was from incorrect to correct. When students changed their answer on developmental items, the most frequent pattern of change was from correct to incorrect. Keeping an initial answer selection was most beneficial on DA items when comparing the two content subjects.

continued on next page

The use of more time and the outcomes associated with GA items might be explained by the fact that the DA content within this course was self-study using only online materials. GA item content was pulled from introductory and clinical lectures, readings, and self-study modules. Although item difficulty was not formally recorded, both content subjects consisted of a mix of first and second order items. Knowing that content for DA items was only coming from one specific resource, students might have been able to better understand what was expected of them, which could have resulted in a lower average item time and less answer changing and/or doubting themselves compared to GA content. There is further research needed on the impact of subject, source material, and item difficulty on answer changing and time.

Limitations

A limitation to this study is that data from only one block exam was analyzed within this first-year medical school student sample. By analyzing data from only one exam, researchers were not able to determine if these outcomes were consistent and standard for this sample across all course exams throughout the semester. Another factor not under consideration was the order of the items within the randomized exams each student received. Because the exam time was capped at 180 minutes, or three hours, the time limit could have affected the presence or lack of a change on an item, especially for the items towards the end of the exam for each specific student. An additional limitation is that the performance quartiles students were assigned to were solely founded upon their performance on the third block exam and not the entire course.

Future Directions

Further research is needed to better understand the impact of specific item attributes on the answering behaviors of first-year medical students. Avenues for future research would be to follow a cohort of first-year medical students throughout an entire course to better understand if their answer changing behaviors are consistent, and likewise, their performance on the exams. Observing the answering behaviors of students within different performance quartiles over time would allow for a more solid understanding of their actions; in turn, this would allow for better supported suggestions to students, especially those who typically perform lower.

Furthermore, because of the differences this study noted in answer changing practices between GA and DA items, future research would likely benefit from breaking down the subject content further into specific anatomical content areas (e.g., leg, thigh, pelvic blood supply, etc.) for all anatomical regions. Some content regions might be more difficult or sensitive for students, and this could influence both their comprehension of the content, their decision to change their answer, and/or the time they spend on an item. Given the results, it would also be beneficial for researchers to observe whether items are first or second order for analysis. Researching the more in-depth item attributes could build upon the findings of this study and provide a stronger understanding of how the item content influences the presence and pattern of answer changing for students at various performance levels.

Additionally, it would be beneficial to measure students' confidence in their answer selections. By determining the confidence of students' initial answer selections and the answer changes they make, researchers would be able to better determine if answer selections and their associated outcomes result from guessing and/or luck or a true understanding and comprehension of the material. This is especially important when considering the differences between students in various performance quartiles.

Conclusion

Educators should hesitate before giving blanket, 'one-size-fits-all' recommendations on answering practices to their students and classrooms, as it could be harmful to specific students. The results of this study support the idea that students who perform at different levels might need to take different approaches to taking an exam and changing answers. In contrast to existing literature, this study showed that those who kept their initial answer selection were the highest performing students and those who logged the most-answer changes performed the lowest. Furthermore, high performing students were most likely to log a single answer change from incorrect to correct, with all other student populations showing a strong tendency to change their answer multiple times and end in an incorrect selection. Based on this study's findings, whether the advice for students is to 'stick with their gut' or 'revisit and rethink,' these claims will likely not be beneficial to all students in the classroom.

The time spent on items that resulted in an incorrect selection were, on average, 30 seconds higher than those that ended with a correct selection. For students, these findings can better guide their answering behaviors in terms of how long they allow themselves to stay on or revisit an item. The pattern of change that took students the longest amount of time was multiple changes ending in a correct response, but the next two longest average pattern times were a single change from correct to incorrect and a single change from incorrect to incorrect. This leads to the conclusion that allowing yourself to rethink and/or revisit an item for long periods of time is not always a beneficial answering behavior and has the potential to lead to negative outcomes. As a result of this study, it seems that advice early in the semester to develop sound metacognitive approaches and build confidence with the material would be more valuable than any exam day advice on answer changing.

continued on next page

About the Authors

Marli Crabtree, MS, is a Clinical Anatomy PhD student at the University of Mississippi Medical Center. She has an MS degree in Anatomy Education from Johns Hopkins School of Medicine. Kenneth L. Thompson, PhD, is an Assistant Professor at the University of Mississippi Medical Center. He teaches statistics and research methodology. Additionally, he serves as the Director of Institutional Research in the Office of the Associate Vice Chancellor for Academic Affairs. Ellen M. Robertson, PhD, OTR/L, is an Assistant Professor in the Department of Physician Assistant Studies at Randolph-Macon College. She teaches gross anatomy, neuroanatomy, and research methods.

Literature Cited

Archer, N. S. & Pippert, R. (1962). Don't change the answer! An exposé of the perennial myth that the first choices are always the correct ones. *The Clearing House, 37*(1), 39-41. https://www.jstor.org/stable/30194626

Bauer, D., Kopp, V., & Fischer, M. R. (2007). Answer changing in multiple choice assessment change that answer when in doubt – and spread the word! *BMC Medical Education, 7,* Article e28. https://doi.org/10.1186/1472-6920-7-28

Benjamin L. T. Jr., Cavell, T. A., & Shallenberger, W. R. (1984). Staying with initial answers on objective tests – Is it a myth? *Teaching of Psychology, 11*(3), 133-141. https://doi.org/10.1177/009862838401100303

Coffey, J. S., Maruca, A. T., Polifroni, E. C., & Snyder, M. (2024). Changing test answers: A scoping review. *Nurse Education Today*, Article e106052. https://doi.org/10.1016/j.nedt.2023.106052

Cox-Davenport, R. A., Haynes, P. B., & Lawson, T. G. (2014). A mixed-methods approach to evaluating student nurses changing answers on multiple choice exams. *Journal of Nursing Education and Practice, 4*(2), 132-139. http://doi.org/10.5430/jnep.v4n2p132

Ferguson, K. J., Kreiter, C. D., Peterson, M. W., Rowat, J. A., & Elliott, S. T. (2002) Is that your final answer? Relationship of changed answers to overall performance on a computer-based medical school course examination. *Teaching and Learning in Medicine, 14*(1), 20-23. https://doi.org/10.1207/S15328015TLM1401_6

Fischer, M. R., Herrmann, S., & Kopp, V. (2005). Answering multiple-choice questions in high-stakes medical examinations. *Medical Education, 39*(9), 890-894. https://doi.org/10.1111/j.1365-2929.2005.02243.x

Friedman, S. J. & Cook, G. L. (1995). Is an examinee's cognitive style related to the impact of answer changing on multiple-choice tests? *The Journal of Experimental Education*, *63*(3), 199-213. https://www.jstor.org/stable/20152451

Geiger, M. A. (1997). An examination of the relationship between answer changing, testwiseness, and examination performance. *The Journal of Experimental Education, 66*(1), 49-60. https://www.jstor.org/stable/20152543

Harvill, L. M. & Davis, G 3rd. (1997). Medical students' reasons for changing answers on multiple-choice tests. *Academic Medicine, 72*(10), S97-S99. https://doi.org/10.1097/00001888-199710001-00033

Liu, O. L., Bridgeman, B., Gu, L., Xu, J., & Kong, N. (2015). Investigation of response changes in the GRA revised general test. *Education and Psychological Measurement*, *75*(6), 1002-1020. https://doi.org/10.1177/0013164415573988

Matthews, C. O. (1929). Erroneous first impressions on objective tests. *Journal of Educational Psychology, 20*(4), 280-286. https://psycnet.apa.org/doi/10.1037/h0071721

McMorris, R. F., DeMers, L. P., & Schwarz, S. P. (1987). Attitudes, behaviors, and reasons for changing responses following answer-changing instruction. *Journal of Educational Measurement, 24*(2), 131-143. https://www.jstor.org/stable/1434691

McNulty, J. A., Sonntag, B., and Sinacore, J. M. (2007). Test-taking behaviors on a multiple-choice exam are associated with performance on the exam and with learning style. *Journal of the International Association of Medical Science Educators, 17*, 52-57. https://medicine.wright.edu/sites/medicine.wright.edu/files/page/attachments/Test-taking_Behaviors_MCQ-Exam.pdf

Merry, J. W., Elenchin, M. K., and Surma, R. N. (2021). Should students change their answers on multiple choice questions? *Advances in Physiology Education*, *45*(1), 182-190. https://doi.org/10.1152/advan.00090.2020

continued on next page

Miller, J., Ruppel, K., Kane, C., Schloemer, B., Power, P., & Owens, H. (2023). Analyzing traditional and accelerated nursing student exam answer changing behaviors: A retrospective study. *Nurse Education Today, 120,* Article e105602. https://doi.org/10.1016/j.nedt.2022.105602

Mueller, D. J., & Wasser, V. (1977). Implications of changing answers on objective test items. *Journal of Educational Measurement, 14*(1), 9-13. https://www.jstor.org/stable/1433850

Ouyang, W., Harik, P., Clauser, B. E., & Paniagua, M. A. (2019). Investigation of answer changes on the USMLE step 2 clinical knowledge examination. *BMC Medical Education, 19,* Article e389. https://doi.org/10.1186/s12909-019-1816-3

Pagni, S. E., Bak, A. G., Eisen, S. E., Murphy, J. L., Finkelman, M. D., & Kugel, G. (2017). The benefit of a switch: Answer-changing on multiple-choice exams by first-year dental students. *Journal of Dental Education, 81*(1), 110-115. https://doi.org/10.1002/j.0022-0337.2017.81.1.tb06253.x

Pettijohn, T. F. II, & Sacco, M. F. (2007). Multiple-choice exam question order influences on student performance, completion time and perceptions. *Journal of Instructional Psychology, 34*(3), 142-149.

Stylianou-Georgiou, A. & Papanastasiou, E. C. (2017). Answer changing in testing situations: The role of metacognition in deciding which answers to review. *Educational Research and Evaluation, 23*(3-4), 102-118. https://doi.org/10.1080/13803611.2017.1390479

Back to TOC