Should students change their answers on multiple choice questions?

Justin W. Merry, Mary Kate Elenchin, and Renee N. Surma

Biology Department, Saint Francis University, Loretto, Pennsylvania

Abstract

Multiple choice exams are ubiquitous, but advice on test-taking strategies varies and is not always well informed by research. This study evaluated the question of whether students benefit or are harmed when they change their initial answers on multiple choice questions in the context of physiology and biology courses. Previously marked examinations were reviewed for eraser marks that indicated answer changes, and the impact of these changes on exam grades was tabulated. In addition, faculty and students were surveyed for their opinions about changing answers. A plurality of faculty (36%) reported a belief that answer changes usually harm student grades, whereas a slim majority of students (51%) believed that answer changing helped their scores ($\chi^2 = 60.52, P < 0.0001$). Empirically, across two exams, students changed their answer from an incorrect answer to a correct one 2.8 times (SD 2.2) compared with 1.0 time (SD 1.4) changing in the negative direction. Therefore, on average, students benefited ($V = 123.5, P < 0.0001$) from answer changing. Furthermore, comparing across two exams in the same course, some students were consistently more likely to change their answers than others (adjusted $R^2 = 0.23, P < 0.0001$), but the impact of changing answers on the first exam provided no prediction of how much a student would benefit from answer changing on the second exam (adjusted $R^2 = -0.004, P = 0.42$). These data support the argument that students should be advised to review and revise responses to exam questions before submitting them.

answer changing; education; multiple choice exams; physiology instruction; test-taking strategy

INTRODUCTION

Multiple choice exams are a pervasive feature of secondary and undergraduate education and are the primary question format for most standardized admissions exams. Although there is substantial literature debating the limitations associated with these types of questions (1, 2), students have little choice but to master this exam format to succeed in many fields of science education. A consequence of this reality is that if educators wish to evaluate students on content knowledge and understanding, as opposed to test-taking ability, they should teach students effective test-taking strategies (3). One common line of advice that we anecdotaly hear at our institution is an admonition against changing one’s initial answer on a multiple choice question: “go with your gut” and “don’t overthink it” seem common refrains from students and faculty alike.

This advice is countered by studies that have examined answer changing on examinations. Past reviews and meta-analyses (4–6) have found that students change their answers from wrong to right significantly more often than they change them from right to wrong. Subsequent work has generally been supportive of this finding and has probed additional nuances regarding answer changing (7–12), although some studies have reached a different conclusion (13, 14). Nevertheless, in general, this literature suggests that students should change their initial responses that they think are incorrect. Even so, there may be reason to think that attitudes toward answer changing, as well as the outcomes of answer changing, may have shifted in the current generation of college students. Under the No Child Left Behind and Every Student Succeeds Acts, children in the United States are subjected to standardized testing involving multiple choice examinations throughout their primary and secondary education (15). Given the impact of test scores on evaluation of districts, elementary teachers are incentivized to not only “teach to the test” but provide specific instruction on multiple choice test-taking strategies (16, 17). This instruction background may result in differences in how today’s college students approach multiple choice exam questions compared with previous generations. Furthermore, interindividual differences among learners may be critically important in analyzing these data (14). Although some prior studies have attempted to use student sex (10), nationality (18), cognitive style (19), ability level (14), anxiety level (20, 21), or other predefined measures of student type to predict success at answer changing, we are not aware of studies that looked for interindividual, repeatable differences among students across multiple exams to evaluate whether certain individuals are inherently more likely, or less likely, to
benefit from answer changing. Finally, there may be nuance that is specifically related to academic discipline. Past studies have focused primarily on courses in education and psychology—reflecting the discipline of the principal investigators—with minimal sampling in the context of physiology education. Of the 39 studies reviewed by Benjamin et al. (4), only 2 included data from the sciences. Subsequent work has also included dental (22), nursing (20, 23, 24), and medical (8, 13) school examinations, but there has been minimal sampling of undergraduate courses in physiology or biology. Despite often relying heavily on multiple choice questions, undergraduate examinations in physiology often require students not just to regurgitate but to respond at higher levels of Bloom’s Taxonomy and Webb’s Depth of Knowledge (3, 25). Question difficulty may impact the effectiveness of answer changing (26), and one of the few studies to find minimal benefit to answer changing was in the context of medical dermatology examinations (13).

For these reasons, we undertook a study to examine answer changing on examinations in introductory biology, human anatomy and physiology, and neuroscience courses. The primary alternative hypotheses tested were 1) answer changing benefits student exam performance and 2) answer changing is harmful to examination performance. Through tracking individual performances across multiple examinations, we also evaluated the hypothesis that some individual students are consistently more likely to be harmed (or to benefit) from answer changing compared with the population as a whole. Our study also included assessments of whether academic rank (freshmen through seniors) or sex differed in answer changing frequency or outcomes. Finally, we conducted a survey to assess opinions regarding answer changing among students and faculty at our institution.

**MATERIALS AND METHODS**

Survey of Student and Faculty Attitudes about Answer Changing

Campus e-mail was used to circulate a link to electronic surveys about answer changing that were administered with Google Forms. Separate surveys were created for students and for faculty, and e-mail distribution lists were employed to deliver the relevant surveys to all students and all faculty. The e-mail was titled “Opinions on Exam Taking,” and the body of the e-mail identified the researchers and explained that we were “seeking opinions about changing answers on multiple choice exams at Saint Francis University,” along with a link to the electronic survey. Saint Francis University had approximately 2,300 students, 130 full-time faculty, and 135 adjunct faculty at the time of the survey. Both surveys asked participants to provide basic demographic information (sex, academic major or academic school) and then report their opinion about whether answer changing helps or hurts grades and whether they (or, for students, their instructors) provide advice to students regarding answer changing (specific wording of questions provided in Table 1). We subsequently classified students into academic school on the basis of their first reported major; we judged this method to be more reliable than asking students to self-report academic school because many students are unaware of this information.

**Review of Exams**

Existing, marked examinations from three courses were utilized in this study: 1) BIOL 111 Introductory Biology: Molecules, Cells, and Animal Physiology; 2) BIOL 205 Human Anatomy & Physiology I; and 3) NEUR 279 Introduction to Neuroscience. All classes were taught by J. W. Merry at Saint Francis University between 2015 and 2017. Within each course, we evaluated two “midterm” examinations for each student: one given approximately one-third of the way through the semester and the second given approximately two-thirds of the way through the semester. The exams were all administered on paper, and students were instructed to circle the best response from the options given. No instruction was given on the days of the examinations about changing initial answers. The examinations each had between 34 and 39 multiple choice questions with 3 to 5 possible answers for each question (mode = 4). No matching or true-false questions were on the exams. Other parts of the exams included fill in the blank, short answer, or sketching questions, which were not considered in this study. All students in these courses had 50 min to complete their exams, and the total number of questions did not vary substantially among the courses: 46 and 49 total questions on the two exams in BIOL 111, 44 and 46 questions on the two exams in BIOL 205, and 47 and 43 questions on the two exams in NEUR 279. Therefore, the students’ available time per question, and thus time available for revision of initial answers, should be similar across the classes.

Before reviewing the exams, we assigned each student a numerical code that we wrote on the first page of the exam booklet, and we then blacked out the name of the student with a Sharpie. In a separate spreadsheet file, we linked each individual student code to the student’s sex, first academic major, and academic rank (based on earned credits at the time of enrollment) but then did not access this spreadsheet again until after all exam data collection was complete. Student names never appeared in our data collection spreadsheet, and demographic information was not connected to exam response data until the spreadsheets were imported and merged in R (see Statistics and Clearances below). Exam batches were evaluated one at a time (e.g., midterm 1 for BIOL 111, then midterm 2 for BIOL 111), and each batch was split haphazardly three ways by the researchers, so that there was never an opportunity during data collection to compare the two exams by the same student. This process concealed student identity and minimized opportunity for the researchers to be mindful of any one student’s first exam vs. second exam performance.

On each examination, we visually searched page by page for eraser marks or “scratch-outs” that indicated a change in answer. Although we pilot the use of light tables to back-illuminate exams, most exam booklets were printed in two-sided format, and so the light tables made exams more difficult to read. We ultimately opted to review exams at a research laboratory benchtop with bright, direct, overhead fluorescent illumination. In each case where there was
Table 1. Survey questions and responses

| Student Survey (n = 318 responses) | Faculty Survey (n = 97 responses) |
|------------------------------------|-----------------------------------|
| How do you identify your sex?      | How do you identify your sex?     |
| 250 (78.6%) [Female]               | 50 (51.5%) [Female]               |
| 66 (20.8%) [Male]                 | 47 (48.5%) [Male]                |
| 1 (0.3%) [Write-in: trans nonbinary] | [Write-in]                     |
| Your major:                        | What is your main school affiliation? |
| 36 (11.4%) [Arts and Letters]     | 35 (37.2%) [Arts and Letters]    |
| 46 (14.6%) [Business]             | 15 (16%) [Business]              |
| 166 (52.7%) [Health Sciences]     | 23 (24.5%) [Health Sciences]     |
| 62 (19.7%) [Sciences]             | 21 (22.3%) [Sciences]            |
| Do you regularly take exams or quizzes with multiple choice questions? | Do you regularly administer exams or quizzes with multiple choice questions? |
| 290 (91.8%) [Yes]                 | 74 (76.3%) [Yes]                 |
| 26 (8.2%) [No]                    | 23 (23.7%) [No]                  |
| When you change your answer on multiple choice exams, you think it will most likely: | When students change their answers on multiple choice exams, I think it will most likely: |
| 162 (51.3%) [Help your score]     | 6 (6.2%) [Help their scores]     |
| 84 (26.6%) [Hurt your score]      | 38 (39.2%) [Hurt their scores]   |
| 24 (7.6%) [Neither help nor hurt your score] | 20 (20.6%) [Neither help nor hurt their scores] |
| 46 (14.6%) [Not sure]             | 33 (34%) [Not sure]              |
| Do your instructors regularly provide advice about changing answers on multiple choice exams? | Do you regularly provide your students with advice, in class or individually, about changing answers on multiple choice exams? |
| 62 (19.6%) [Yes]                  | 24 (24.7%) [Yes]                 |
| 255 (80.4%) [No]                  | 73 (75.3%) [No]                  |
| Which of these best represents the advice given by your instructors at SFU about changing answers on multiple choice exams? | Which of these best represents the advice you give your students about changing their answers on multiple choice exams? |
| 2 (0.6%) [Instructors recommend that students consider changing their initial answers] | 3 (3.1%) [I recommend that students consider changing their initial answers.] |
| 113 (35.6%) [Instructors recommend that students do not change their initial answers.] | 18 (18.8%) [I recommend that students do not consider changing their initial answers.] |
| 35 (11%) [The recommendation varies.] | 13 (13.5%) [My recommendation differs from student to student.] |
| 167 (52.7%) [Instructors do not give advice about answer changing on multiple choice exams.] | 62 (64.6%) [I do not give my students advice about answer changing on multiple choice exams.] |
| Any additional comments? [Write-in] | Any additional comments? [Write-in] |

Questions with no responses were ignored for the purpose of calculating percentages. Available responses are listed in brackets, along with the number of respondents, and the percentage of respondents, who provided that answer.

Evidence of a change in answer, the change was classified as one of the following: “right to wrong,” “wrong to right,” or “wrong to wrong.” Given that these were existing, previously marked exams, incorrect answers were indicated by red pen circling the correct answer. Students’ answers were usually marked in pencil or blue/black ink. For each student exam, we counted the number of answer changes in each category and recorded those totals in a spreadsheet along with the student code for later connection to demographic information. In the rare cases where a researcher was uncertain on whether marks indicated an answer change, or how to classify a question, the exam was passed around the table for the opinion of the other two individuals. We initially had two researchers evaluate each exam, but, as was found in prior studies of this kind (4), interobserver reliability was extremely high. Therefore, the individual student exams recorded in our data set were each processed by one of the three study authors.

Statistics and Clearances

Data analysis was conducted in R (27). Differences in survey results by students and faculty were calculated with chi-square tests of independence on contingency tables summarizing the survey data. For analysis of answer changing data from our review of marked exams, we initially used a Wilcoxon signed-rank test to check for differences in the number of changes from wrong to right and right to wrong for each student (summed across both exams). This simple test was followed by a linear mixed-model approach using the lme4 package (28), again using the answer changes summed across both exams for each student. Our null model included random intercepts for each Student. The following three fixed effects were then added to the model, and log-likelihood ratios were calculated to test whether they improved the model’s fit: Change Type (wrong to right, right to wrong, or wrong to wrong), Rank (freshman, sophomore, junior, or senior), and Sex (female or male). We then sequentially considered models that added several of these variables, following an iterative, forward-selection selection procedure. Interactions between fixed effects were also considered, although none provided a significant improvement in fit. We did not test for differences across the three classes, because class was confounded with academic rank: introductory biology was primarily comprised of freshmen, anatomy and physiology consisted of sophomores and juniors, and neuroscience included sophomores and juniors but was also the only course that included seniors (Table 2). In all analyses of answer changing, the count data were square root-transformed so that the variances were independent of the means (29).

Finally, to check for individual tendencies in answer changing, we used ordinary least-squares regressions to
Table 2. Demographics of the three courses from which marked exams were evaluated in this study, organized by academic rank and sex

| Course                                | Freshmen | Sophomores | Juniors | Seniors |
|----------------------------------------|----------|------------|---------|---------|
| B I O L 1 1 I Intro to Biology (n = 2 2 students) | Female 1 5 | 2 | 1 | |
|                                        | Male 4 | | | |
| B I O L 2 0 5 Human Anatomy & Physiology (n = 2 5 students) | Female 8 | 1 0 | | |
|                                        | Male 3 | 4 | | |
| NEUR 2 7 9 Intro to Neuroscience (n = 3 1 students) | Female 4 | 1 1 | 8 | |
|                                        | Male 1 | 2 | 1 | 5 |

Values shown are the number of students in each category. Note that table includes only those individuals used in subsequent data analysis.

compare the number of answer changes on exam 1 versus exam 2 as well as (in a second analysis) the effect of answer changing on exam 1 versus exam 2. If the frequency or the effect of answer changing differs consistently between students, then students’ actions on the first exam should predict their actions on the second exam. To evaluate the effect of answer changing, we calculated the net questions improved for each exam (the number of wrong-to-right changes minus the number of right-to-wrong changes). A student who made two changes from wrong to right but one change from right to wrong would have a net questions improved score of “+1.”

The methods for the electronic survey and the use of pre-existing, anonymous, graded examinations were approved by the Institutional Review Board at Saint Francis University before the start of the study.

RESULTS

Survey of Student and Faculty Attitudes about Answer Changing

Three hundred eighteen students and 97 faculty responded to their respective surveys, which represents approximately 14% of the student body and 37% of faculty (full-time and adjunct faculty) at Saint Francis University. One student response was removed from the data set because it included obviously inappropriate/joking responses. Both faculty and students were well represented across the institution’s four academic schools: Arts and Letters, Business, Health Sciences, and Sciences (Table 1). The faculty data set was evenly divided between sexes (51.5% female, 48.5% male), whereas the student responses were biased toward women [78.9% female, 20.8% male, 0.3% (1 student) trans nonbinary]. This bias in student sex ratio reflects the disparities in the Saint Francis student body, which is 63% female and 37% male (undergraduate population, Fall 2017 enrollment). We did not ask students to report academic rank in our survey, which was conducted before the review of exams.

Strikingly, there was a significant difference in student and faculty attitudes toward answer changing. A majority of students felt that when they changed their answers on multiple choice exams they helped their score (51% helped; 26% hurt; 8% neither helped nor hurt; 15% not sure). In contrast, a plurality of faculty felt that answer changing hurt student exam scores (36% hurt); only 6 of 97 (9%) faculty respondents felt that answer changing helped exam scores (Fig. 1, Table 1; \( \chi^2 \) test of independence, \( \chi^2 = 60.52, P < 0.0001 \)). Among students, there was no indication of differences in these opinions based on sex (\( \chi^2 \) test of independence within students only, comparing sex and survey answers, \( \chi^2 = 7.15, P = 0.31 \)) or academic school based on first major (\( \chi^2 = 8.23, P = 0.51 \)). Similarly, faculty also did not differ in opinion on the utility of changing answers based on sex (within faculty only, \( \chi^2 = 5.75, P = 0.12 \)) or academic school affiliation (\( \chi^2 = 12.62, P = 0.18 \)).

Students and faculty agreed on whether advice was given by faculty: a majority indicated that they received (students, 52%) or gave (faculty, 64%) no instruction on whether to change answers. When advice was given, students either were instructed not to change answers (students: 36%, faculty: 19%) or the recommendation varied (students: 11%, faculty: 13%). Very few indicated that faculty coach students to consider changing their initial answers (students: 2%, faculty: 3%).

Student Answer Changing Outcomes

We evaluated 157 exams taken by 79 students across three classes (21 students in B I O L 1 1 I Intro to Biology, 26 students in B I O L 2 0 5 Human Anatomy & Physiology I, and 31 students in NEUR 2 7 9 Intro to Neuroscience). One student withdrew from the course between the first and second exams in Human Anatomy & Physiology I and therefore was removed from the analysis; Table 2 presents demographic information on the remaining 156 exams and 78 students. As with the survey data, the exam data were dominated by women (59 female vs. 19 male). Across the data set, there was fairly even representation across academic rank (19 freshmen, 19 sophomores, 27 juniors, and 13 seniors). However, because of the nature of the courses, enrollment was dominated by individuals from the health sciences: 65 Health Sciences majors, 7 Sciences majors, 2 Arts and Letters majors, and 4 undeclared majors.

Overall, we found that, on average, students changed their answers from wrong to right 2.8 times (SD 2.2) per student across both exams compared with just 1.0 time (SD 1.4) from right to wrong, which constituted a significant difference based on a Wilcoxon signed-rank test (\( V = 123.5, P < 0.0001 \), using square root-transformed values; Fig. 2A). On average, changes from a wrong answer to another wrong answer occurred 1.0 time (SD 1.4) per student.

We used a linear mixed-model approach to further explore the relationship between the type of answer change, sex, and academic rank of individuals. This analysis again found significant differences between the types of answer changes (log-likelihood ratio for comparison of null model vs. model including answer change type, again using square root-transformed values: \( \chi^2 = 78.2, P < 0.0001 \)); there were significantly more wrong-to-right changes than right-to-wrong changes (Tukey post hoc test: \( t = -8.49, P < 0.0001 \)) and wrong-to-wrong changes (\( t = 8.84, P < 0.0001 \); Fig. 2A). The frequency of answer changing also differed based on academic rank (null model vs. model including rank: \( \chi^2 = 8.2, P = 0.042 \)); freshmen [6.95 changes (SD 4.71) per student]...
changed their answers significantly more than seniors [2.92 changes (SD 2.14) per student; Tukey post hoc test between freshmen and seniors: \( t = 2.59, P = 0.05 \); other comparisons were not significantly different; Fig. 2B]. The consequence of changing answers did not differ across academic rank, however. Within each rank, students averaged many more changes from wrong to right than from right to wrong [freshmen: 3.5 (SD 2.8) wrong to right vs. 1.5 (SD 2.0) right to wrong; sophomores: 2.7 (SD 2.4) wrong to right vs. 1.3 (SD 1.3) right to wrong; juniors: 2.7 (SD 1.6) wrong to right vs. 0.7 (SD 1.0) right to wrong; seniors: 2.1 (SD 1.6) wrong to right vs. 0.5 (SD 0.9) right to wrong]. Furthermore, there was no indication of an interaction between answer changing type and academic rank (\( \chi^2 = 7.18, P = 0.30 \)). Therefore, students who changed answers benefited more often than they were hurt, regardless of academic rank. Finally, there were no detectable differences between sexes in answer changing (\( \chi^2 = 0.17, P = 0.68 \)).

To evaluate whether there were repeatable differences among individuals, we compared individual student answer changing on the first versus the second exam. Students differed consistently in how frequently they changed answers: the number of answer changes on exam 1 was a significant predictor of that student’s number of changes on exam 2 (least-squares regression on square root-transformed values: \( t = 4.88, P < 0.0001 \); adjusted \( R^2 = 0.23 \); Fig. 3A). However,
this association did not result in measurable, repeatable benefits in exam scores for some students over others. Within each exam, there was a significant, positive correlation between the number of answers a student changed and their consequential number of questions improved on that exam (Spearman rank-order correlation: \( \rho = 0.52, P < 0.0001 \) on exam 1; \( \rho = 0.35, P = 0.0017 \) on exam 2). Nevertheless, net questions improved by answer changing on exam 1 did not predict net questions improved on exam 2 (\( t = -0.81, P = 0.42; \) adjusted \( R^2 = -0.004; \) Fig. 3A). In other words, there was no evidence that some students consistently were harmed or helped by answer changing decisions in ways that were different from other students. Instead, all students benefited by an average of about one net question improvement on exam 2 (\( y \)-intercept estimate = 1.1 ± 0.2 SE, range of \(-4 \) to \(+6 \) question improvement, \( t = 5.2, P < 0.0001 \)), regardless of the outcome of their answer changes on exam 1. Among 78 students in our data set, only 1 negatively affected their scores on both exams that semester through answer changes.

## DISCUSSION

Our survey results found striking differences in opinions about answer changing between students and faculty. Whereas a majority of students felt that their answer changes helped their scores, a plurality of faculty believed that answer changing hurt student scores. A similar number of faculty were unsure, but only a few faculty believed that answer changing helped student performances. Our data on the actual consequences of answer changing on exams were consistent with prior studies (4, 5): overall, answer changing is beneficial by approximately a 3-to-1 ratio. Furthermore, we found no indication that certain students were more likely to be harmed or helped than others by their answer changing decisions. Therefore, the opinions advocated by faculty, although consistent with prior surveys of this kind (4), seem to be unfounded.

Qualitatively, comments from faculty indicated that many believed that the benefit of answer changing most likely differs from student to student. For example, one faculty member wrote “Overall I’d say that the percentage of students who change from right to wrong is roughly equal to the percentage that change from wrong to right. I look at the individual trend with each student that comes to review their exams and base my advice on their tendencies.” Others emphasized a cautious approach: “I recommend that they do not consider changing unless upon reviewing the question they are very sure that the initial choice is incorrect.” Another wrote: “I recommend students only change answers when they are convinced the answer was wrong.”

Several faculty noted that the tendency to believe that changing answers is harmful could be the result of cognitive bias: “People tend to remember the answer that was changed when it was wrong, but not if it was correct.” Indeed, this same argument has been made in past investigations of this topic (4, 30). Instructors reviewing exams with a student will typically focus on questions that the student got wrong, and therefore are far more likely to notice examples where students changed their answer from right to wrong. The result may be that faculty are reinforced to believe this “instinct fallacy” (31).

It is notable, therefore, that the surveyed students responded with such comparative favorability regarding answer changing. This contrasts strongly with past published surveys. For example, a review of prior studies by Benjamin...
et al. (4) reported that “approximately three out of every four of these students felt answer changes would lower their score.” By contrast, our survey found that a (slim) majority (51%) of students believed that answer changes had helped their scores on exams. Therefore, it is possible that the message of past research on this topic is being communicated to the current generation of college students. Doing so might help them. Stylianou-Georgiou and Papanastasiou (26) found that students with a bias against answer changing performed slightly worse on exams. Furthermore, Bauer and colleagues (7) found that students were more likely to change answers when informed of the benefits of doing so, although it resulted in an undetectable improvement in test scores. Other work, on the other hand, found no change in behavior (32).

It is unclear where the students in our survey might have received such information. It is not coming from their current faculty: on the survey, students reported that if faculty gave advice at all, they typically recommended against answer changes (36% vs. 0.6% recommended to change answers). It is possible that past teachers, guidance counselors, or other test preparation materials informed their view. That said, of the 40 students who added free-form comments to our survey, only 1 mentioned advice from someone other than a professor. In that case, a high school teacher had recommended against changing answers.

There is the possibility that the way that we asked the questions about outcomes of answer changing could have impacted the responses on our surveys. Faculty were asked to respond to “When students change their answers on multiple choice exams, I think it will most likely...” Therefore, faculty were asked to generalize across students. Students were asked a more personal question: “When you change your answer on multiple choice exams, you think it will most likely...” Students who already guard themselves against answer changing might respond that answer changes help them, because they are careful to only change answers when they think it likely that the answer they initially made was wrong. Nevertheless, the narrowed perspective of the question to students also mirrors how we assessed actual student performances: we could only measure cases in which students did change a question, not those cases in which the student considered changing a question but opted not to do so (14). Faculty may have been asked to generalize but still were asked to focus on situations in which answer changes occur, rather than any instance where answer changes are considered.

Our survey results represented roughly 14% of the student body and 37% of all faculty (full time and adjunct), based on Saint Francis University institutional totals. The survey was designed to be short and was advertised to all students and faculty via a single e-mail distribution (no reminders were sent). Our response rates were within the ranges reported in some other studies that use e-mailed survey invitations to college campuses [e.g., Dykema et al. (33)] reported response rates of 8.6% initially and up to 19.4% after a second reminder e-mail; Muñoz-Leiva et al. (34) reported 16% initially and up to 48% after 5 reminders]. Furthermore, our data set captured students and faculty from all four academic schools on campus (Arts and Letters, Business, Health Sciences, and Sciences), and the student data set, like our campus population, was heavily skewed toward females (78.6% female vs. 20.8% male; Table 1). Nevertheless, there remains the possibility of sampling bias. Given that the focus of the survey was noted clearly in the recruitment e-mail, those without strong opinions or interest in the topic would be expected to have lower participation rates. Therefore, the proportion of individuals in our sample responding with intermediate/uncertain responses (e.g., “neither help nor hurt” or “not sure” on the question of how answer changing impacts exam scores) may be lower than the theoretical population mean. We also cannot rule out the possibility that response bias in the survey affected the proportion of responses indicating that that answer changes helped or hurt on exams, although we have no indications that this is the case, nor do we have a way to predict the direction of such bias.

The question of why students ultimately decide to change answers (or not) has been probed by researchers. Past studies that asked students to reflect on why they changed their answers (23, 26, 35) reported a variety of reasons: rethinking questions, rereading questions and gaining new understanding, clerical mistakes, and new insights gained from other questions later in an exam. More difficult questions would be expected to lead to more of these kinds of circumstances. Not surprisingly, Stylianou-Georgiou and Papanastasiou (26) found that more difficult test items, where students also felt the least confident, elicited more answer changes. They also found that these specific increases in answer changing unfortunately led to more right-to-wrong or wrong-to-wrong changes (not wrong-to-right changes). However, other work has reached the opposite conclusion: Couchman et al. (31) found that answer changes were most likely to be successful when a student’s in-the-moment confidence in their initial answer is low, whereas changes when students initially felt confident in their answers were less likely to be successful. Student reflective (post hoc) assessments of confidence in initial answers, however, provided no predictive power. Therefore, although our data and previous data show that student answer changes are generally beneficial when they occur, there is still debate on whether students should adjust their willingness to change answers based on question difficulty or the students’ perceived confidence in their answers.

The data reported in our study indicate that students experienced net benefits when they changed initial answers on multiple choice exams. Furthermore, although students differed in how often they changed their answers, there was no evidence that some students were more likely to benefit or be harmed by answer changing than others. There was, however, an inverse relationship between the number of answer changes and experience in college (i.e., academic rank). Although academic rank was confounded with the course selections (e.g., mostly freshmen in introductory biology; Table 2), we see no evidence that these differences were caused by differences in the courses or exam content. The time allotted (50 min), the total number of questions on the exams (43–49), and the types of questions (similar ratio of multiple choice, fill in the blank, and short answer questions, all written by the same instructor) did not vary substantially among classes or in a way that would explain the differences seen across academic rank. For example, the introductory biology course had the most total questions on its exams (46...
and 49 on exams 1 and 2, respectively, vs. 43–47 in other classes; see MATERIALS AND METHODS), and yet the freshmen in that course showed the largest number of answer changes in our data set.

Higham and Gerrard (9) found that answer changes were most likely to successfully correct problems when correcting errors due to speeded responding and less effective when correcting errors based on confusion or lack of understanding. Furthermore, they found that when students experienced negative past experiences with answer changing, those students were less likely to change answers moving forward. Anxiety may also play a role: George et al. (20) found that students changed their answers in a fashion that correlated with student anxiety level: more anxious students had a higher percentage of changed answers, relative to the number of exams, than less anxious students. It is possible that older, more experienced students 1) had learned to avoid rushing through multiple choice exams and therefore had less need to correct errors due to fast responding; 2) had learned to avoid changing answers through anecdotal negative experiences and consequently reduced their tendency to change their answers; and/or 3) were less anxious than freshmen and therefore changed their answers less often.

Although multiple choice exams delivered on paper have long been a staple of secondary and higher education, many courses, both online and face to face, have moved their examinations into an online testing environment. For example, at our institution it is commonplace for face-to-face classes to administer multiple choice examinations in the classroom via our online learning management system. This change does allow for automatic scoring but may negatively affect the opportunity for students to reconsider and revise their answers. Many online testing systems do not provide the opportunity for students to mark a question so that they may later revisit it, for example, and they do not permit students to annotate their thoughts about questions in the margins. Even worse, some systems display only one question at a time and, in some cases, prevent students from returning to previous questions after answering. Although these latter modifications are motivated by a desire to prevent cheating, we argue that they deprive students of the opportunity to reflect upon their work, learn from that reflection, and ultimately demonstrate their learning through answer changes that improve exam outcomes. Therefore, we urge faculty to be cautious about adopting online testing procedures, particularly if they restrict the opportunity to revise answers.

Given the still-pervasive opinion among faculty, at least at our institution, that changing initial answers harms students, we argue that it is important that the results of studies like this one are communicated to educators. In addition to guiding instruction and encouraging learning, a primary goal of testing in higher education should be the assessment of student learning. To achieve that goal, students should be provisioned with evidence-based advice on test-taking strategies so that their knowledge of the subject, rather than test-taking ability, dictates their scores. Our advice to students, based on the present study as well as past research, is straightforward: “Review your answers before turning in your test. If you think your initial answer is wrong, and another response is correct, change your answer!”

**ACKNOWLEDGMENTS**

We thank all of the student and faculty participants in the survey for their time. Our thanks also go to Dr. Stephanie Ivory of Saint Francis University, and three anonymous reviewers, who provided helpful comments on previous versions of this manuscript.

**GRANTS**

This study was supported by an undergraduate research grant by the Saint Francis University Biology Department to M.K.E.

**DISCLOSURES**

No conflicts of interest, financial or otherwise, are declared by the authors.

**AUTHOR CONTRIBUTIONS**

J.W.M., M.K.E., and R.N.S. conceived and designed research; J.W.M., M.K.E., and R.N.S. performed experiments; J.W.M., M.K.E., and R.N.S. interpreted results of experiments; J.W.M. prepared figures; J.W.M. drafted manuscript; J.W.M., M.K.E., and R.N.S. edited and revised manuscript; J.W.M., M.K.E., and R.N.S. approved final version of manuscript.

**REFERENCES**

1. Stanger-Hall KF. Multiple-choice exams: an obstacle for higher-level thinking in introductory science classes. *CBE Life Sci Educ* 11: 294–306, 2012. doi:10.1187/cbe.11-11-0100.
2. Tiemeier AM, Stacy ZA, Burke JM. Using multiple choice questions written at various Bloom’s taxonomy levels to evaluate student performance across a therapeutics sequence. *Innov Pharm* 2: 5, 20f. doi:10.24926/IP.V2I2.224.
3. McKenzie WJ, Svinicki M. Teaching Tips: Strategies, Research, and Theory for College and University Teachers (12th ed). New York: Houghton Mifflin Company, 2006.
4. Benjamin LT Jr, Cavell TA, Shallenberger W III. Staying with initial answers on objective tests: is it a myth? *Teach Psychol* 11: 133–141, 1984. doi:10.1177/009862838401100303.
5. Mueller DJ, Wasser V. Implications of changing answers on objective test items. *J Educ Meas* 14: 9–13, 1977. doi:10.1111/j.1745-3984.1977.tb00023.x.
6. Waddell DL, Blankenship JC. Answer changing: a meta-analysis of the prevalence and patterns. *J Contin Educ Nurs* 25: 155–158, 1994.
7. Bauer D, Kopp V, Fischer MR. Answer changing in multiple choice assessment change that answer when in doubt—and spread the word. *BMC Med Educ* 7, 28, 2007. doi:10.1186/1472-6920-7-28.
8. Fischer MR, Herrmann S, Kopp V. Answering multiple-choice questions in high-stakes medical examinations. *Med Educ* 39: 890–894, 2005. doi:10.1111/j.1365-2929.2005.02243.x.
9. Higham PA, Gerrard C. Not all errors are created equal: metacognition and changing answers on multiple-choice tests. *Can J Exp Psychol* 59: 28–34, 2005. doi:10.1037/h0087457.
10. Liu OL, Bridgeman B, Gu L, Xu J, Kong N. Investigation of response changes in the GRE Revised General Test. *Educ Psychol Meas* 75: 1002–1020, 2015. doi:10.1177/0013164415573988.
11. Tiemann G. An Investigation of Answer Changing on a Large-Scale Computer-Based Educational Assessment (PhD dissertation). Lawrence, KS: University of Kansas, 2015. http://hdl.handle.net/1808/19513.
12. Vispoel WP. Reviewing and changing answers on computer-adaptive and self-adaptive vocabulary tests. *J Educ Meas* 35: 328–345, 1998. doi:10.1111/j.1745-3984.1998.tb00542.x.
13. Noorbala MT, Mohammadi S. A survey on the habit to change the answers in multiple choice questions (MCQ) exams: does the examinee benefit? J Pak Assoc Dermatologists 21: 253–259, 2011.
14. van der Linden WJ, Jeon M, Ferrara S. A paradox in the study of the benefits of test-item review. J Educ Meas 48: 380–398, 2011. doi:10.1111/j.1745-3984.2011.00151.x.
15. McGuinn P. From No Child Left Behind to the Every Student Succeeds Act: federalism and the education legacy of the Obama administration. Publius 46: 392–416, 2016. doi:10.1093/publius/pwj014.
16. Menken K. Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. Biling Res J 30: 521–546, 2006. doi:10.1080/15235882.2006.10162888.
17. Ryan RM, Weinstein N. Undermining quality teaching and learning: a self-determination theory perspective on high-stakes testing. Theory Res Educ 7: 224–233, 2009. doi:10.1177/1477878509104327.
18. Milia LD. Benefitting from multiple-choice exams: the positive impact of answer switching. Educ Psychol (Lond) 27: 607–615, 2007. doi:10.1080/01443410701309142.
19. Friedman SJ, Cook GL. Is an examinee’s cognitive style related to the impact of answer changing on multiple-choice tests? J Exp Educ 63: 199–213, 1995. doi:10.1080/00220973.1995.9943809.
20. George TP, Muller MA, Bartz JD. A mixed-methods study of prelicensure nursing students changing answers on multiple choice examinations. J Nurs Educ 55: 220–223, 2016. doi:10.3928/01484834-20160316-07.
21. Payne BD. The relationship of test anxiety and answer-changing behavior: an analysis by race and sex. Meas Eval Guidance 16: 205–210, 1984. doi:10.1080/00256307.1984.12022358.
22. Pagni SE, Bak AG, Eisen SE, Murphy JL, Finkelman MD, Kugel G. The benefit of a switch: answer-changing on multiple-choice exams by first-year dental students. J Dent Educ 81: 110–115, 2017. doi:10.1002/j.0022-0337.2017.811606253.x.
23. Gaskins S, Dunn L, Forte L, Wood F, Riley P. Student perceptions of changing answers on multiple choice examinations. J Nurs Educ 35: 88–90, 1996.
24. Jordan L, Johnson D. The relationship between changing answers and performance on multiple-choice nursing examinations. J Nurs Educ 29: 337–340, 1990.
25. Webb NL. Criteria for alignment of expectations and assessments in mathematics and science education. Research Monograph No. 6 (Online). Washington, DC: Council of Chief State School Officers. https://eric.ed.gov/?id=ED414305 [18 May 2020].
26. Stylianou-Georgiou A, Pananastasiou EC. Answer changing in testing situations: the role of metacognition in deciding which answers to review. Educ Res Eval 23: 102–118, 2017. doi:10.1080/13803611.2017.1390479.
27. R Core Team. R: A Language and Environment for Statistical Computing (Online). Vienna, Austria: R Foundation for Statistical Computing, 2020. https://www.R-project.org/.
28. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 67: 1–48, 2015. doi:10.18637/jss.v067.i01.
29. Sokal RR, Rohlf FJ. Biometry: the Principles and Practice of Statistics in Biological Research. New York: W. H. Freeman and Co., 1995.
30. Bath JA. Answer-changing behavior on objective examinations. J Educ Res 61: 105–107, 1967. doi:10.1080/00220671.1967.10883602.
31. Couchman JJ, Miller NE, Zmuda SJ, Feather K, Schwartzmeyer T. The instinct fallacy: the metacognition of answering and revising during college exams. Metacogn Learn 11: 171–185, 2016. doi:10.1007/s11409-015-9140-8.
32. Prinsell CP, Ramsey PH, Ramsey PP. Score gains, attitudes, and behavior changes due to answer-changing instruction. J Educ Meas 31: 327–337, 1994. doi:10.1111/j.1745-3984.1994.tb00450.x.
33. Dykema J, Stevenson J, Klein L, Kim Y, Day B. Effects of e-mailed versus mailed invitations and incentives on response rates, data quality, and costs in a web survey of university faculty. Soc Sci Comput Rev 31: 359–370, 2013. doi:10.1177/0894439312465254.
34. Munoz-Leiva F, Sánchez-Fernández J, Montoro-Ríos F, Ibáñez-Zapata JA. Improving the response rate and quality in Web-based surveys through the personalization and frequency of reminder mailings. Qual Quant 44: 1037–1052, 2010. doi:10.1007/s11135-009-9256-5.
35. Schwarz SP, McMorris RF, DeMers LP. Reasons for changing answers: an evaluation using personal interviews. J Educ Meas 28: 163–171, 1991. doi:10.1111/j.1745-3984.1991.tb00351.x.