

Test Corrections Appear To Benefit Lower-Achieving Students in an Introduction to Biology Major Course: Results of a Single-Site, One-Semester Study

Kyeorda Kemp^a

^aDepartment of Foundational Medical Studies, Oakland University William Beaumont School of Medicine, Rochester, Michigan, USA

Student-performed assessment correction is a well-established practice in the field of teaching and learning. This practice engages students in the feedback process and promotes active learning, which may be particularly important when serving underrepresented minority students. There is a dearth of research into the role of assessment correction in student learning outcomes (SLOs) in science courses, particularly at minority-serving institutions. Students at a Native American-serving, primarily undergraduate rural institution were allowed to perform test corrections for each of the three exams given during the term in an undergraduate introductory biology course. Students received the graded test back and were given 1 week to answer the following questions: what was your original answer, why did you choose that answer/ what were you thinking at the time, why is the answer wrong, what is the correct answer, why is it the correct answer, and what is your source of reference if different from the textbook? The students were administered a comprehensive final exam at the end of the course that determined the number of SLOs passed. A Pearson correlation and a bivariate regression analysis were performed to determine if the number of test corrections performed (TC) during the term influenced the number of SLOs passed for all students, lower-achieving students, and higher-achieving students in the study. The TC correlated with, and could predict, SLOs passed for lower-achieving students only. This preliminary analysis suggests that performing test corrections may benefit lower-achieving students.

KEYWORDS test corrections, student learning outcomes, biology major course, undergraduate, minority-serving institution

INTRODUCTION

Success in science, technology, engineering, and math (STEM)-related fields is dependent not only on what a student knows but also in part on the student's ability to learn from their error. Error can be used to promote student knowledge of material, and error generation can improve memory of the correct answer (1). However, learners require more than a simple statement of right or wrong to understand why they have erred. Previous studies indicate that receiving feedback in the form of right or wrong is not

as useful as being given the correct answer or having a chance to review the material (2–5).

Feedback has traditionally been seen through a lens of transmission, where the authority/expert tells the learner how they are deficient. This view of feedback does not incorporate the student into the feedback exercise. Another understanding of feedback is “a process through which learners make sense of information from various sources and use it to enhance their work or learning strategies” (6, p 1315). This means that for feedback to be effective, students must think critically about the material. The instructor can tell a student that they are deficient in their knowledge, but the student must engage with the material, evaluate their knowledge, and obtain the information needed to bolster their knowledge. The students are no longer passive observers but are active participants in their learning process. Errors made by students provide us with the opportunity to engage learners in the feedback process.

Giving students the opportunity to explore errors and correct mistakes can improve their learning (7, 8). Moreover, this can help promote active learning in the classroom environment. This may prove to be particularly important for schools with large underrepresented

Citation Kemp K. 2021. Test corrections appear to benefit lower-achieving students in an introduction to biology major course: results of a single-site, one-semester study. *J Microbiol Biol Educ* 22:e00122-21. <https://doi.org/10.1128/jmbe.00122-21>.

Address correspondence to Department of Foundational Medical Studies, Oakland University William Beaumont School of Medicine, Rochester, Michigan, USA. E-mail: kyeordakemp@oakland.edu.
Received: 1 December 2020, Accepted: 4 June 2021,
Published: 30 July 2021

minority and economically disadvantaged populations. A 2019 study found that active learning opportunities increased academic achievement and belongingness in students attending a school with large underrepresented minority and economically disadvantaged populations (9).

The literature on the association of metacognition and student success is extensive and has been reviewed in numerous publications (10–15). Lower-achieving students (defined as students who score lower than the course median) have reduced metacognition or understanding about how they learn (11, 16, 17), and metacognition is promoted when students engage in activities that allow reflection upon their learning, knowledge, and study skills (18). Assessment corrections may promote the use of metacognitive strategies to enhance learning in lower-achieving students if students are asked to apply retrospective analysis (19).

Nicol and Macfarlane-Dick propose seven principles of good feedback practice based on a review of formative assessments that promote self-regulation. They contend that good feedback practice should provide opportunities for students to self-assess, engage the teacher and learner in a dialog, and provide opportunities for students to use feedback to improve their performance (20). Allowing students to correct tests can be used to provide effective feedback, as defined above, and promote metacognition. Test corrections that incorporate reflection promote student use of metacognition strategies to enhance their learning because students are asked to apply retrospective analysis to their exams (19). Test corrections can promote self-assessment by requiring students to reflect on their thought processes during the assessment and why they came to a particular conclusion. Test corrections can also increase dialog between the learner and instructor and provide an opportunity for faculty to improve their instruction by requiring a student to explain their thought process and why they reached a conclusion. Moreover, the instructor can determine if a concept was confusing or not covered appropriately. Finally, test corrections allow summative assessments to be used as a tool to measure learning. Black and Wiliam explain that “the giving of marks and the grading function are overemphasized, while the giving of useful advice and the learning function are underemphasized” in the classroom. They further explain that this approach places students in competition with each other (21). Test corrections change the function of the test, as its primary purpose is no longer determining competence or rank. This may prove to be particularly important when working with underrepresented minority populations, first-generation college students, or women engaged in STEM courses, as these populations may underperform on high-stakes exams (22, 23).

Many of the studies exploring student-performed assessment corrections in higher education have been done in non-STEM fields; the few studies performed within the STEM fields focused only on the quality of the assessment corrections, student academic growth, content retention or, have focused on how students feel about the process of self-correction (24–29). This retrospective, exploratory study was conducted to

determine if student participation in test corrections can influence the number of student learning outcomes (SLOs) passed in a freshman-level introductory biology course for biology majors at a minority-serving institution located in a rural setting. A hypothesis was generated: students who participate in test corrections will possess a different pass rate for SLOs compared to students who do not participate in test corrections. A prediction was made: student participation in the feedback process through the act of performing test corrections would result in increased numbers of SLOs passed. The following questions were developed to address this prediction:

- Does the average number of test points received (AT) during the term influence the number of SLOs passed?
- Does preparedness influence the number of SLOs passed?
- Does the number of test corrections performed (TC) during the term influence the number of SLOs passed for all students, lower-achieving students, and higher-achieving students in the study?

METHODS

Human subject research

This study was designated exempt and conducted as approved by the Institutional Review Board (IRB) at Northeastern State University (The Role of Self-Assessment Corrections in Student Learning [IRB number 19-093M]).

Course description and participants. (i) Study site and population

This is a retrospective, exploratory study of existing data collected over the spring semester of 2015 in two sections of a freshman-level introductory biology course at a regional college. The institution is a Native American-serving, nontribal, primarily undergraduate institution located in a rural area. The student body is predominately female (62%), in-state resident (90%), and composed of traditional and nontraditional students. The demographics of the students enrolled in the course can be found in Table 1. While there were 74 students enrolled at the time of the final exam, only 69 students took the final exam. Existing data related to the course from a convenience sample of the 69 students who took the final exam were included in the study as the SLOs were assessed using the final exam.

(ii) Course description

The course was required for the biology major and designed to meet the needs of students pursuing a biology degree; however, the course was also required for a number of other majors and could be used to satisfy the general education requirements for the university. The course covered the following: chemistry and physics of life and the central dogma of molecular biology, which mapped to SLOs 1 and 2 (exam 1); cells

KEMP: TEST CORRECTIONS BENEFIT LOWER-ACHIEVING STUDENTS

TABLE 1
Demographic data for the students enrolled in the course

Demographic	% (no.) of students
Gender	
Female	56.8 (42)
Male	43.2 (32)
Major	
Biology	36.5 (27)
Undeclared	32.4 (24)
Other ^a	31.1 (23)
Race/ethnicity	
American Indian or Alaska Native	20.3 (15)
White	37.8 (28)
2 or more races	29.7 (22)
Other ^b	12.2 (9)

^aOther includes students from majors with fewer than 10 students each. Majors represented are chemistry, criminal justice, health and human performance, human and family sciences, music, psychology—general, science education, medical laboratory science, and environmental, health, and safety management.

^bOther includes students from categories with fewer than 10 students each. Categories represented are Hispanic/Latino, Asian, black or African American, and unknown.

and cell division, Mendelian and non-Mendelian inheritance, mechanisms and evidence of evolution, phylogenies, and genes and evolution, which mapped to SLOs 3, 4, and 8 (exam 2); species and speciation, history of life on earth focusing on the evolution of species, and prokaryotic diversity, which mapped to

SLOs 5, 6, and 7 (exam 3); and the comprehensive final exam, which covered eukaryotic diversity and the above-described topics, which mapped to all SLOs (Table 2). The course was delivered face to face via lecture; however, students participated in discussion, group work, and think-pair-share during the lectures. In addition, students had weekly homework sets and biweekly online formative quizzes where students were allowed multiple attempts to provide the correct answer. The catalog description, course purpose, and SLOs can be found in Table 2.

Data sources. (i) Test corrections and timeline

Students were given tests back with in 2 weeks after taking the exam and allowed 1 week to find the correct answers to the exam questions (Fig. 1). The tests were a combination of multiple-choice, matching, and short-answer questions. The students had to answer the following questions when completing test corrections:

- What was your original answer?
- Why did you choose that answer/what were you thinking at the time?
- Why is the answer wrong?
- What is the correct answer?
- Why is it the correct answer?
- What is your source of reference if different from the textbook?

Students received half of the original credit back for each answer if they were correct and provided an appropriate rationale. Credit was awarded to encourage students to perform test corrections; awarding credit improves participation in assessment corrections (30). Students who submitted an

TABLE 2
Information regarding the course

Course information
Catalog description of course
3 h; an introduction to the origins of living organisms and the mechanisms of evolution that gave rise to the current diversity of species; includes coverage of the origins and characteristics of major groups in the 3 domains of living organisms
Course purpose
This course is designed for biology majors; this lecture course will provide students information about the basic principles and mechanisms of evolution as well as an overview of the past and present diversity of life on earth
Student learning outcomes for the course
1. Describe the structure and function of DNA and its importance to evolution
2. Describe transcription, translation, and other biological processes relevant to evolution
3. Explain the relationship between genes, gene expression, and traits
4. List and describe the stages of meiosis and relate their importance to evolution
5. List and describe the various mechanisms of evolutionary change in natural populations
6. Explain how the process of evolution has led to the development of modern life forms through the process of speciation and differentiation
7. List and describe the names and characteristics of major groups of living organisms on earth, including the 3 domains of life and the 4 major groups of eukaryotes
8. Interpret evolutionary information in the form of a phylogenetic tree or cladogram

KEMP: TEST CORRECTIONS BENEFIT LOWER-ACHIEVING STUDENTS

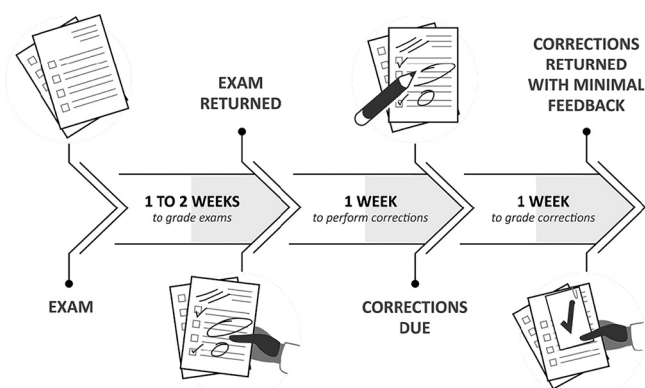


FIG 1. Timeline for test corrections. The students were given 1 week to perform corrections. The corrections were made within 1 week and returned with minimal feedback from the instructor.

incorrect answer again were marked incorrect and told to correct their work but for no credit. Students were told why their reasoning was flawed in instances where the instructor felt that the student required additional support based on the student's submitted explanation and were encouraged to meet with the instructor for help.

(ii) Assessment measures

Students took a final exam that contained questions that tested whether the student achieved competency regarding the course's eight SLOs at the end of the term. Competency is defined in this study as the minimum knowledge and/or skills required of a student and is determined by the program, while SLOs are statements that describe what a student should be able to do by the end of the course and were assessed on the final exam. Students who pass all SLOs are considered competent. The questions that mapped to the SLOs were designed by the author and another instructor. In most cases, a score of 60% was considered competent since 60% was the minimum score needed to pass the course with a D for all students enrolled in the course. For example, for SLOs 1 through 7, there were five questions, of which students had to answer three questions

correctly to achieve an average score of 60% and be considered competent. Unfortunately, for SLO 8, one of the questions was removed due to an error in the question stem. For this reason, there were only four questions in total; therefore, students were required to answer three of the four questions correctly and achieve an average score of 75% to be considered competent. Students were given an optional preassessment test to assess their general knowledge of the material to be covered in the course, and 62 of the students completed the assessment (see Appendix SI in the supplemental material); this was used as a measurement of preparedness for the course to answer the question of whether preparedness for the course could influence the role of test corrections in the ability of students to achieve SLOs.

Data analysis

Descriptive statistics were compiled (Tables 3 to 5), and analysis of the following variables occurred: the number of test corrections (TC) that students chose to perform during the term (0, 1, 2, and 3), the number of SLOs passed (0 to 8), the average number of test points received (AT) on the three exams given during the term, and the preassessment test points received (PT). Student's unpaired *t* test assumption of unequal variance or an independent-sample Mann-Whitney U test was performed to determine if the distributions of the variables in the individual sections were similar (Supplemental Tables 1 and 2 in Appendix SII). The tests were chosen based on whether or not the data were normally distributed. The sections were combined into one data set after determining that the distributions and variances of the individual sections were similar. The Shapiro-Wilk test was performed on the data set, in addition to the creation of histograms, to determine normality (Supplemental Table 3 and Fig. S1 in Appendix SII) (31, 32). Statistical tests were chosen as appropriate based on these analyses (parametric versus nonparametric). Analysis of variance (ANOVA) (single factor) was performed to determine if the AT received for the three tests given during the term differed between the groups of students who completed 0, 1, 2, or 3 test corrections. A Kruskal-Wallis test was performed to determine if there was a difference between the SLOs passed and PT

TABLE 3

Descriptive statistics for student learning outcomes passed for students who performed zero, one, two, and three test corrections and for all students^a

No. of test corrections	No. of students	No. of SLOs passed					
		Median	Minimum	Maximum	Mode	Mean	SD
0	21	4.00	0.00	8.00	2.00	4.48	2.46
1	16	4.50	0.00	8.00	8.00	5.13	2.39
2	13	6.00	1.00	8.00	8.00	5.85	2.15
3	19	5.00	3.00	8.00	5.00	5.63	1.64
Total	69	5.00	0.00	8.00	8.00	5.20	2.21

^aThe number of students in each group is recorded. There is no statistically significant difference for student learning outcomes (SLOs) between the groups of students who performed 0, 1, 2, and 3 test corrections [$H(3) = 3.582$; $P = 0.310$].

TABLE 4

Descriptive statistics for preassessment test scores for students who performed zero, one, two, and three test corrections and for all students^a

No. of test corrections	No. of students	PT				
		Median	Minimum	Maximum	Mean	SD
0	16	10.13	8.00	13.75	10.73	1.76
1	15	11.00	6.50	13.25	10.77	1.90
2	13	10.50	5.25	13.75	9.69	2.48
3	18	10.63	4.75	13.25	10.56	1.96
Total	62	10.50	4.75	13.75	10.47	2.01

^aThe number of students in each group is recorded. There is no statistically significant difference for preassessment test scores (PT) between the groups of students who performed 0, 1, 2, and 3 test corrections [$H(3) = 1.267; P = 0.735$].

for students based on the number of test corrections performed. Pearson’s correlation and bivariate regression analyses were performed between SLOs and the variables AT and number of test corrections (TC) for the complete data set. The data set was also sorted based on AT and analyzed as described above based on lower (lower-achieving students [TC-LA]) and upper (higher-achieving students [TC-HA]) medians, and Pearson’s correlation and bivariate regression analyses were performed between SLOs and TC. Confidence intervals were set at 95%, and a two-tailed analysis was performed for all tests. All analysis was done using Excel or SPSS, with the exception of the independent-sample Mann-Whitney U test for the PT variable and the Kruskal-Wallis test. Online calculators were used for these two analyses (<https://www.socscistatistics.com/tests/mannwhitney/default2.aspx> and <https://www.socscistatistics.com/tests/kruskal/default.aspx>).

RESULTS

Does the average number of test points received during the term influence the number of SLOs passed?

The mean, median, and mode values for the average number of test points received (AT) were determined for the combined total of students in the study and for

students based on the number of test corrections performed (TC). There was no statistically significant difference for AT between the groups based on the number of test corrections performed (Table 3). The scatterplot in Fig. 2 indicates that there was a strong positive linear relationship between the AT and SLOs passed. A bivariate regression was calculated to predict SLOs passed based on AT and showed a statistically significant relationship between AT and SLOs passed ($P < 0.001$) (Table 6).

Does preparedness influence the number of SLOs passed?

The mean, median, and mode values for PT were determined for the combined total of students in the study and for students based on the number of test corrections performed. There was no statistically significant difference for PT between the groups based on the number of test corrections performed (Table 4). The scatterplot in Fig. 3 shows that there was a weak, negative linear relationship between PT and SLOs passed. A bivariate regression was calculated to predict SLOs passed based on PT and showed a statistically nonsignificant relationship between PT and SLOs passed ($P = 0.937$) (Table 6).

TABLE 5

Descriptive statistics for average test points obtained per exam during the term for students who performed zero, one, two, and three test corrections and for all students^a

No. of corrections	No. of students	AT				
		Median	Minimum	Maximum	Mean	SD
0	21	35.71	13.00	62.40	37.35	12.19
1	16	39.30	19.24	61.88	38.30	12.24
2	13	41.60	22.36	56.68	41.47	10.09
3	19	43.16	29.47	57.55	43.85	7.79
Total	69	40.21	13.00	62.40	40.14	10.87

^aThe exams were out of 75 points. There is no statistically significant difference for average test points obtained (AT) between the groups of students who performed 0, 2, 3, and 3 test corrections ($F_{3,65} = 1.447; P = 0.237$).

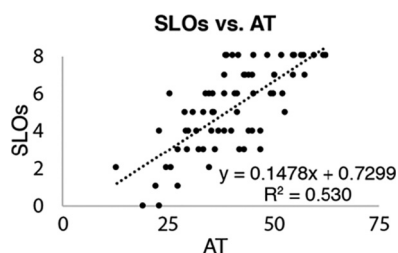


FIG 2. Scatterplot showing the relationship between SLOs passed and the average test points received (AT) on the test during the term. The scatterplot showing the relationship between SLOs and AT was generated for all students ($n = 69$).

Does the number of test corrections performed during the term influence the number of SLOs passed for all students, lower-achieving students, and higher-achieving students in the study?

The SLO mean, median, and mode were determined for the combined total of students in the study and for students based on the number of test corrections performed. There was no statistically significant difference for SLOs between the groups based on the number of test corrections performed (Table 5). The scatterplot in Fig. 4A shows that there was a weak, positive linear relationship between TC and SLOs passed. A bivariate regression was calculated to predict SLOs passed based on TC for all participants enrolled in the study and showed a statistically nonsignificant relationship between TC and SLOs passed ($P = 0.066$) (Table 6). The data set was sorted based on AT and analyzed based on the lower (lower-achieving students [TC-LA]) and upper (higher-achieving students [TC-HA]) medians. The scatterplots in Fig. 4B and C show that there was a strong, positive linear relationship between TC and SLOs passed for lower-achieving students but a weak, positive linear relationship between TC and SLOs passed for higher-achieving students. A bivariate regression was calculated to predict SLOs passed based on TC for lower-achieving students and showed a statistically significant relationship between TC-LA and SLOs passed ($P = 0.015$) (Table 6). This was not so for TC-HA and SLOs ($P = 0.126$) (Table 6).

DISCUSSION

Research indicates that minority and female students have reduced self-efficacy, have higher anxiety, and benefit from active learning approaches (33, 34). Active learning opportunities appear to increase academic achievement and belongingness in students attending schools with large underrepresented minority populations (9). Test corrections promote an active learning environment by creating an opportunity for students to reflect on their learning and why they erred, engage the student in the feedback process, allow the instructor to determine if a concept is communicated clearly, and may provide an opportunity for students

to increase their metacognitive skills. While there are many benefits to test corrections, the exercise is labor-intensive for the instructor, and some find it prohibitively burdensome (24). Therefore, an exploratory study using existing data from an introductory biology course was initiated in order to determine if performing test corrections increased the SLO pass rate at a minority-serving institution located in a rural setting. This study found that the number of test corrections performed correlated with and could predict the number of SLOs passed for lower-achieving students only, while AT correlated with and could predict the number of SLOs passed for all students.

Previous studies exploring the role of test corrections on assessment scores in STEM courses found them to be beneficial for students at lower academic levels. Mynlieff et al. used ACT scores to sort students by academic level and found that students with low ACT scores (17 to 23) benefited more than those with higher scores from test corrections in an introductory biology course (25). Unlike this study, they found that test corrections positively benefited the performance of all students. The discrepancy between these studies may be due to the previous study assessing student performance 2 weeks after versus at the end of the term, as was done in this study. Alternatively, sample size differences may play a role (35). In addition, studies exploring the impact of student performance of quiz corrections in a physics course (28, 29) found that interventions that require students to self-assess their problems appeared to reduce the difference in scores between lower- and higher-achieving students.

Lower-achieving students have reduced metacognitive monitoring (11, 18), and metacognition is promoted in lower-achieving students when they engage in activities that allow them to reflect upon their learning, knowledge, and study skills (16). Moreover, analysis of test strategy has a positive effect on exam preparation and performance (36). Mynlieff et al. found similar gains in the intervention groups of students with low ACT scores who participated in test corrections and of those who participated in small-group discussions of incorrectly answered questions. However, students with high ACT scores benefited from test corrections only (25). Students who are higher achievers may already be participating in activities that promote metacognition; therefore, they do not benefit from

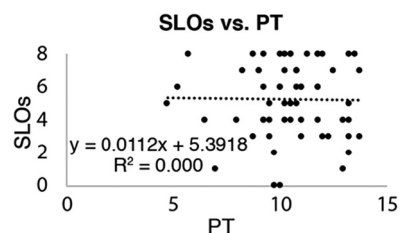


FIG 3. Scatterplot showing the relationship between SLOs passed and the preassessment test score (PT). The scatterplot showing the relationship between SLOs and PT was generated for all students who took the preassessment test ($n = 62$).

KEMP: TEST CORRECTIONS BENEFIT LOWER-ACHIEVING STUDENTS

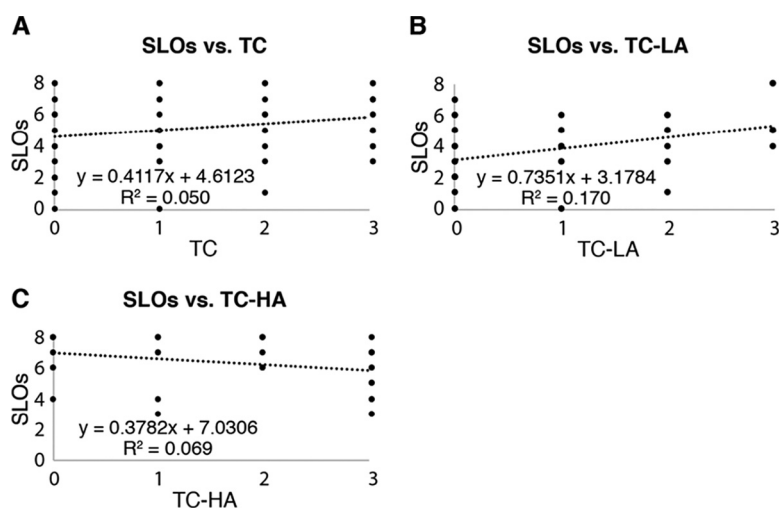


FIG 4. Scatterplots showing the relationship between SLOs passed and test corrections completed (TC). Scatterplots showing the relationship between SLOs and TC were generated for all students ($n = 69$) (A), low-achieving students ($n = 34$) (B), and high-achieving students ($n = 35$) (C).

exam discussions (25). If true, this enhanced metacognitive monitoring could mask the effect of test corrections in the higher-achieving cohort analyzed in this study. Interestingly, there was a weak, positive linear relationship between AT and TC, and a simple linear regression indicated that AT could predict TC (see Fig. S1 in the supplemental materials of Appendix SIII). This indicates that higher-achieving students in this study were more likely to participate in an activity that promotes metacognition. While AT correlated with SLOs passed, PT did not. This is interpreted to mean that a student may come in less prepared, but if they gain the skills required to be effective in their learning, they have an increased probability of success. Training students to self-assess errors, by reflecting on their knowledge and learning, helps students determine what they need to do to achieve (21).

In summary, the act of performing test corrections may be beneficial for lower-achieving students, as TC correlated with and was predictive of SLOs passed for lower-achieving students only. While AT correlated with and was predictive of SLOs passed for all students, PT did not correlate with SLOs passed. It is predicted that a student's ability to adapt

metacognitive strategies may be the strongest indicator of the number of SLOs passed. It is proposed that efforts be made to identify students with poor study skills and provide them with resources to help them develop self-assessment, correcting skills, and metacognitive strategies.

Limitations

The sample size for this study is small and could potentially mask the impact of test corrections on SLOs (35). Moreover, this study was performed on one cohort of students at a regional school, and these students may not be representative of the general college population. However, this study may be informative for universities and colleges that serve minority and rural populations. In addition, while the SLO assessment questions were decided by the author and another faculty member based on the SLOs and the material covered during the course, the SLO assessment questions were not validated by an outside source.

This study did not explore the quality of the test corrections, the difficulty of the questions corrected, or the number of questions corrected per student for each test correction

TABLE 6
Bivariate regression analyses performed^a

Parameter	Value for group				
	TC	TC-HA	TC-LA	PT	AT
Coefficient (SE)	0.412 (0.220)	0.264 (0.241)	0.413* (0.2487)	-0.011 (0.142)	0.148**** (0.017)
Correlation coefficient	0.223	-0.378	0.735	0.10	0.728
R^2	0.050	0.069	0.170	0.000	0.530
Adjusted R^2	0.354	0.041	0.145	-0.017	0.523
No. of observations	69	35	34	62	69

^aThe dependent variables are shown in the left column, and the independent variables are shown at the top. *, P value of <0.05 ; ****, P value of <0.0001 .

assignment. This limited analysis may mask the effect of performing test corrections; however, the results of this exploratory study suggest that the act of performing test corrections may be beneficial for lower-achieving students.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, DOCX file, 0.4 MB.

ACKNOWLEDGMENTS

I thank Demetri Plessas, Jason Wasserman, and Dan Gildner for their help regarding the statistical analysis. I thank Stephen Loftus, Mark Paulissen, Judith Venuti, Nadia Wiczorkowski, and Barbara Joyce for their help editing the manuscript. I thank Ableser and Amanda Hess for advice regarding publishing this study.

The Oakland University William Beaumont School of Medicine supported this study.

I declare no conflicts of interest.

REFERENCES

1. Metcalfe J. 2017. Learning from errors. *Annu Rev Psychol* 68:465–489. <https://doi.org/10.1146/annurev-psych-010416-044022>.
2. Butler AC, Karpicke JD, Roediger HL. 2007. The effect of type and timing of feedback on learning from multiple-choice tests. *J Exp Psychol Appl* 13:273–281. <https://doi.org/10.1037/1076-898X.13.4.273>.
3. Metcalfe J, Kornell N. 2007. Principles of cognitive science in education: the effects of generation, errors, and feedback. *Psychon Bull Rev* 14:225–229. <https://doi.org/10.3758/bf03194056>.
4. Butler AC, Roediger HL. 2008. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem Cognit* 36:604–616. <https://doi.org/10.3758/mc.36.3.604>.
5. Anderson RC, Kulhavy RW, Andre T. 1972. Conditions under which feedback facilitates learning from programmed lessons. *J Educ Psychol* 63:186–188. <https://doi.org/10.1037/h0032653>.
6. Carless D, Boud D. 2018. The development of student feedback literacy: enabling uptake of feedback. *Assess Eval High Educ* 43:1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>.
7. Booth JL, Lange KE, Koedinger KR, Newton KJ. 2013. Using example problems to improve student learning in algebra: differentiating between correct and incorrect examples. *Learn Instr* 25:24–34. <https://doi.org/10.1016/j.learninstruc.2012.11.002>.
8. Siegler RS. 1995. How does change occur: a microgenetic study of number conservation. *Cogn Psychol* 28:225–273. <https://doi.org/10.1006/cogp.1995.1006>.
9. Wilton M, Gonzalez-Niño E, McPartlan P, Ternner Z, Christoffersen RE, Rothman JH. 2019. Improving academic performance, belonging, and retention through increasing structure of an introductory biology course. *CBE Life Sci Educ* 18:ar53. <https://doi.org/10.1187/cbe.18-08-0155>.
10. Coutinho S. 2007. The relationship between goals, metacognition, and academic success. *Educate* 7:39–47.
11. DiFrancesca D, Nietfeld JL, Cao L. 2016. A comparison of high and low achieving students on self-regulated learning variables. *Learn Individ Differ* 45:228–236. <https://doi.org/10.1016/j.lindif.2015.11.010>.
12. Hrbáčková K, Hladík J, Vávrová S. 2012. The relationship between locus of control, metacognition, and academic success. *Procedia Soc Behav Sci* 69:1805–1811. <https://doi.org/10.1016/j.sbspro.2012.12.130>.
13. Kaplan M, Silver N, LaVaque-Manty D, Meizlish D (ed). 2013. Using reflection and metacognition to improve student learning: across the disciplines, across the academy. Stylus Publishing, Sterling, VA.
14. Kramarski B, Mevarech ZR, Arami M. 2002. The effects of metacognitive instruction on solving mathematical authentic tasks. *Educ Stud Math* 49:225–250. <https://doi.org/10.1023/A:1016282811724>.
15. White BY, Frederiksen JR. 1998. Inquiry, modeling, and metacognition: making science accessible to all students. *Cogn Instr* 16:3–118. https://doi.org/10.1207/s1532690xci1601_2.
16. Isaacson RM, Fujita F. 2006. Metacognitive knowledge monitoring and self-regulated learning: academic success and reflections on learning. *J Scholarsh Teach Learn* 6:39–55.
17. Shahapur NP. 2018. A study of metacognitive awareness of high achievers, average and low achievers of central school students. *Int J Res Soc Sci* 8:347–361.
18. Dang NV, Chiang JC, Brown HM, McDonald KK. 2018. Curricular activities that promote metacognitive skills impact lower-performing students in an introductory biology course. *J Microbiol Biol Educ* 19:19.1.5. <https://doi.org/10.1128/jmbe.v19i1.1324>.
19. Tanner KD. 2012. Promoting student metacognition. *CBE Life Sci Educ* 11:113–120. <https://doi.org/10.1187/cbe.12-03-0033>.
20. Nicol DJ, Macfarlane-Dick D. 2006. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud High Educ* 31:199–218. <https://doi.org/10.1080/03075070600572090>.
21. Black P, Wiliam D. 2010. Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan* 92:81–90. <https://doi.org/10.1177/003172171009200119>.
22. Ballen CJ, Salehi S, Cotner S. 2017. Exams disadvantage women in introductory biology. *PLoS One* 12:e0186419. <https://doi.org/10.1371/journal.pone.0186419>.
23. Madaus G, Clarke M. 2001. The adverse impact of high stakes testing on minority students: evidence from 100 years of test data, p 85–106. In Orfield G, Kornhaber M (ed), *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. Century Foundation Press, New York, NY.
24. Henderson C, Harper KA. 2009. Quiz corrections: improving learning by encouraging students to reflect on their mistakes. *Phys Teach* 47:581–586. <https://doi.org/10.1119/1.3264589>.
25. Mynlieff M, Manogaran AL, St Maurice M, Eddinger TJ. 2014. Writing assignments with a metacognitive component enhance learning in a large introductory biology course. *CBE Life Sci Educ* 13:311–321. <https://doi.org/10.1187/cbe.13-05-0097>.

26. Posner M. 2011. The impact of a proficiency-based assessment and reassessment of learning outcomes system on student achievement and attitudes. *Stat Educ Res J* 10:3–14.
27. Ramírez Balderas I, Guillén Cuamatzi PM. 2018. Self and peer correction to improve college students' writing skills. *Profile Issues Teach Prof Dev* 20:179–194. <https://doi.org/10.15446/profile.v20n2.67095>.
28. Yerushalmi E, Cohen E, Mason A, Singh C. 2012. What do students do when asked to diagnose their mistakes? Does it help them? I. An atypical quiz context. *Phys Rev Phys Educ Res* 8:e020109.
29. Yerushalmi E, Cohen E, Mason A, Singh C. 2012. What do students do when asked to diagnose their mistakes? Does it help them? II. A more typical quiz context. *Phys Rev Phys Educ Res* 8:e020110.
30. Brown BR, Mason A, Singh C. 2016. Improving performance in quantum mechanics with explicit incentives to correct mistakes. *Phys Rev Phys Educ Res* 12:e010121.
31. Ghasemi A, Zahediasl S. 2012. Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab* 10:486–489. <https://doi.org/10.5812/ijem.3505>.
32. Mohd Razali N, Bee Wah Y. 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *J Stat Model Anal* 2:21–33.
33. Harris RB, Mack MR, Bryant J, Theobald EJ, Freeman S. 2020. Reducing achievement gaps in undergraduate general chemistry could lift underrepresented students into a “hyperpersistent zone.” *Sci Adv* 6:eaz5687. <https://doi.org/10.1126/sciadv.aaz5687>.
34. Hood S, Barrickman N, Djerdjian N, Farr M, Gerrits RJ, Lawford H, Magner S, Ott B, Ross K, Roychowdury H, Page O, Stowe S, Jensen M, Hull K. 2020. Some believe, not all achieve: the role of active learning practices in anxiety and academic self-efficacy in first-generation college students. *J Microbiol Biol Educ* 21:21.1.19. <https://doi.org/10.1128/jmbe.v21i1.2075>.
35. Kim HY. 2015. Statistical notes for clinical researchers: type I and type II errors in statistical decision. *Restor Dent Endod* 40:249–252. <https://doi.org/10.5395/rde.2015.40.3.249>.
36. Favero TG, Hendricks N. 2016. Student exam analysis (debriefing) promotes positive changes in exam preparation and learning. *Adv Physiol Educ* 40:323–328. <https://doi.org/10.1152/advan.00060.2016>.