Measuring the Active Learning Gains of Students In Undergraduate Biology Labs Using Peer Review

By Joseph Bantel

A thesis submitted to the Department of Biological Sciences, University of Manitoba,

in partial fulfilment of the requirements for the course

BIOL 4100 (Honours Thesis)

for the degree of

Bachelor of Science (Honours)

Abstract

The peer-review process allows students to engage in the valuable exercise of critically thinking about how they generate and judge the validity of an idea. This is a cornerstone of the scientific community and a core component of biological sciences. Some students in undergraduate studies routinely underperform in the traditional educational experience and benefit greatly from active learning. In a first-year introductory biological science course at the University of Manitoba, I studied the population by using an online peer-review platform called Peerceptiv to study the standardized learning gains of different demographic groups of students to determine if any subgroups experienced a disproportional benefit in this learning exercise compared to any others. I hypothesized that peer-to-peer feedback on scientific graphing would raise the measurable graded assignment scores of URM students in undergraduate biological science, and therefore predicted that there would be no significant difference between URM and non-URM students' graded performance after peer-to-peer input. Students in certain groups did have higher learning gains over the course of the semester when averaged however, these results were not significantly different from the rest of the population. Overall, the reception of the exercise was very positive in student perceptions with 79% of respondents saying that it helped them understand the concepts they learned about. The strongest response was from students saying that getting reviews of their work was the most useful, with 87%. Ultimately, this study shows that students learn from the peer review process and have learning gains associated with it however, additional testing is necessary to definitively say which groups may or may not benefit the most from this activity.

Acknowledgements

To Cassandra Debets, thank you for the wonderful learning experience in teaching as a TA and studying biology education as my advisor. I learned so much from this project.

To Dr. Joy Stacey, thank you for introducing me to science and critical thinking in a university setting from first year biology to an honours student six years later. Your advice over the years has been invaluable and I have learned so much from you about what kind of educator I want to be.

To Tamarah Kelly, your advice on this project was invaluable and with your encouragement I look forward to doing more research in education in the future.

To Crystal, Aella, and Ransom, you suffered so much over the last 6 years while I worked hard to finish this degree, let alone this paper. I will never be able to repay your sacrifices to the extent that I should. Thank you for your support.

Thank you to all those who said, “you’ve got this”, you were right, but the phrase makes me sick now from hearing it so many times.

No thank you to the community who used the phrase “Do your own research”. I did it, my therapist will be billing you directly for the trauma you caused.

Table of Contents

List of Tables:

List of Figures:

Introduction

Learning includes a large umbrella of study, including both biological and psychological components (Deslauriers et al., 2019; Piaget et al., 1952). The overall biological mechanistic theories are primarily concerned with the neurophysiological processes. The prefrontal cortex is responsible for the identification and initial signalling of visual stimuli for learning. The dorsal region first recognizes and learns the new information, and then the recurring identification of the learned model is repeated throughout a more widely distributed area. (Meyers et al., 2012). Other areas of study are more interested in the psychological theories of learning. The psychological theories include four main schools of approach for study: behaviourist, cognitive, constructivist, and social (Helou et al., 2018; Maddox et al., 2011; Orsmond et al., 2006). Social learning, in particular, has an interactive component that differs from the other psychological components of learning as defined by the Social Component of Learning Theory by Wenger (Illeris, K. 2008. Pg 209-232).

Social learning sets itself apart from them by being an interactive experience between two or more individuals. Social learning in this peer-to-peer space is an aspect of active learning and involves the demonstration and imitation of skills (Bandura, 1977; Illeris, K. 2008, Pg 209-232; Smith et al., 2009). The increased engagement between differing demographic groups of students has a positive correlation with an increase in self-reported critical thinking and analysis (Strayhorn 2009). Social learning is important for active learning to be more effective and impactful in and out of the classroom. The peer-to-peer space especially requires increased social interactions.

Active learning involves frequent tests of application, notetaking, critical thinking, and peer engagement (Freeman et al. 2007; Deslauriers et al. 2011; Fraser et al. 2014; Hake 1998). It is a vocal, involved, dynamic, and sometimes unpredictable process (Bonwell & Eison, 1991). In the active learning environment, the feedback that students receive becomes a conversation that is more than simply correcting what they may or may not have gotten wrong; it becomes an integral part of the normalization of learning together and supporting each other. In-class open discussion clickers, and short answer in-class quizzes in a digital format, are examples of active learning shown by Freeman (et al. 2007) where students gain understanding by engaging with questions that stimulate discussion and the recall of previously heard or read knowledge. Experiments and activities that ask for the application of knowledge assist further and in course experiments led by Deslauriers, Fraser, and Hake, results show that when students apply the knowledge from notes taken and use critical thinking with peers to problem solve they learn better than simply when using the socratic lecture method (Deslauriers et al. 2011; Fraser et al. 2014; Hake 1998).

Though the active learning approach benefits all students, it does not benefit all students equally (Theobold et al., 2020; Strayhorn, 2009). This is particularly true for Underrepresented minority students (URM). In a review and meta-analysis of studies that measured student performance (15 studies on student examination totalling 9,238 students and student failure rates from 26 studies including 44,606 students), the achievement gap was decreased between URM and non-URM students by 33% on identical or similar examinations when using active learning (-0.62 SDs in traditional lecture vs –0.42% SDs in active learning lectures) (Theobold et al., 2020). In North America, this group consists of a wide range of people who are underrepresented in their given field and more specifically in STEM fields in comparison to the general population. These groups of students experience higher attrition rates in finishing a degree in six years than others in the cohort. Completion rates for students in the USA fall to around 52% for Asian Americans, 43% for Caucasians, 22% for African Americans, 29% for Latinx, and 25% for Indigenous (National Academy Press, 2016). Typically, students are asked to self-declare racial or gender categories to describe themselves and their experience (Ballen et al. 2017). For my definition of URM, any student not Caucasian or Asian (excluding the Indian subcontinent) and in the biologically male category will be considered a URM student. In this case, biological male applies to those who are called male at birth and do not later express a different gender than assigned at birth. An example of a URM student at the University of Manitoba with multiple factors for status would be a first-generation university student of Indigenous descent, part of the LGBTQIA2+ community, and a biological female (Ballen et al., 2017; Estrada et al., 2016).

Peer review can be an extension of active learning outside of the classroom and used in student assessment. Using peer review in scientific writing involves writing your paper and then having those in your field review the information and give feedback about the quality of your writing (Topping, 2009). In practice, students write out a piece of writing and grade or edit each other's writing. The student then analyzes the feedback, and improvements are made. As a tool in university, it mirrors much of the peer review process in scientific publications. Students who engage in this process gain valuable insight and possibly more frequent personal feedback about their writing than when they turn in an assignment to an instructor who may have limited resources as the student-to-student ratio is far better than an instructor with 120 students (Huisman et al., 2019).

Evidence shows that students, especially URM students, gain learning experience from peer feedback despite their lack of experience in scientific writing (Biango-Daniels & Sarvary, 2021). Being a relatively young field, quantitative data with stable controls in large data sets to determine the effectiveness of peer review has needed to be more robust compared to other fields. Prior to the creation and acquisition of the Peerceptiv platform by the University of Manitoba, made it difficult to measure and track the improvements of the high number of students that would provide accurate analysis. The use of peer review as an assessment tool is quite young and its effectiveness has not been measured extensively in prior studies as without Peerceptiv comparisons become more difficult to ascertain and measure. (Debets, Unpublished). A meta-analysis of the literature on peer-to-peer feedback resulted in better student performance in scientific writing assignments compared to no feedback or self-assessment (Huisman et al., 2019). When the teacher gave feedback, the results were similar to writing improvements in the group that peer-reviewed each other's work. The quality of writing did not improve significantly with feedback from multiple peer sources and was just as effective if only one person gave genuine peer-to-peer feedback. At the time of the analysis, there was still little methodologically sound research, and the presented examples had sample sizes too low to be considered reliably significant (Huisman et al., 2019).

Rationale/Hypothesis/Predictions

The goal of this study is to measure the difference in the variation that peer-to-peer feedback has on biological undergraduate students in a diverse student population. From observations at the University of Manitoba, the gap between the success rate of URM students in final overall grades and the rest of the student population is large enough to warrant the intervention of active learning (Debets, Unpublished). Theobald (et al. 2020) demonstrated reduced learning gaps and higher performance gains for underrepresented minority students. Therefore, I hypothesize that peer-to-peer feedback on scientific graphing will raise the measurable graded assignment scores of URM students in undergraduate biological science, and therefore predict that there will be no significant difference between URM and non-URM students' graded performance after peer-to-peer input. I predict that larger learning gains will be found among URM students, and those who fit multiple categories of URM status will show the most increase in learning gains. It is expected that second-generation university students, White and Asian (excluding the Indian subcontinent), and biological, straight males will have the lowest learning gains with peer-to-peer learning. I expect the learning gains expressed by the population will increase over the three assignments during the semester and that the gap in learning gains between URM and non-URM students will decrease.

Methods

Course Description

The first-year biology labs (BIOL 1020 B01 - B10) at the University of Manitoba are introductory-level courses for students from a broad range of disciplines. This is often one of the first courses taken in the Faculty of Science and comes with a steep learning curve that provides a stark contrast between high school and university-level studies as students must be more self motivated and the intellectual difficulty increases. This lab-based portion of the course is a prerequisite for many other biology-related courses and is linked with a lecture component. It is a direct prerequisite for BIOL 1030 to continue in the faculty of science majoring in Biology. Students in this course learn basic principles of biology, which include bioenergetics, cell division, genetics, evolution, and cell biology.

Graphing in the BIOL 1020 labs involves instruction during labs 1 and 2 on how to create, interpret, and present statistical data in the biological context and the appropriate formatting for that information. Students learn about the importance of peer-to-peer feedback in university-level studies to inform and improve performance on assignments. This assists in decreasing outliers of extreme improvement from the draft to the final submission of the assignments.

Assignment Description

Graphing in the sciences is a skill that many students struggle to understand appropriately and provides an opportunity to evaluate a skill needed across scientific disciplines (Gardner et al., 2024). There were six lab assignments throughout the course which were all based on graphing scientific data collected during the lab period. Assignments (2, 3, 6) were submitted to a platform online called Peerceptiv where students engaged in the process of peer review by submitting graphing assignments for initial review by two peers who then graded based on a rubric provided (Appendix Part B).

Demographic and Perception Survey

The data collection for surveying students was carried out according to a protocol approved by the University of Manitoba human ethics research board. Students were advised of its need for completion on UMLearn as part of the grade for the course (value 1%). Student names will not be used in the analysis of the data, and students were given the choice to consent to the study after completion of the questionnaire (See Appendix part A) research was carried out according to a protocol approved by the University of Manitoba human ethics research board."

Statistical Methods

I removed students who obtained a zero on assignments 2,3, and 6 for both the Peerceptiv and the UmLearn submission to eliminate students who, for whatever reason, did not engage with the process. For overall grades for assignments not analyzed in the comparative analysis of the effectiveness of Labs 2,3, and 6, I left in scores of 0 or blanks, which in the group already selected likely represent students who finished the course but simply did not finish one of the assignments not included in this study. These scores were kept to keep an overall sense of the student's grade for later analysis. The remaining total for my population is 787 students.

Students' racial data were grouped when the declared racial identity was geographically close to another regional area for the purposes of testing the groups. If students had two different racial identities, they were categorized as “Mixed” if the declared identities were not amalgamated into larger groups (Table 1). In the cohort, I differentiated between racial declaration of indigenous heritage and indigenous status, as some individuals identify as indigenous, however, they also declared multiple racial profiles. The category of indigenous includes any individual who declared indigenous heritage when it is part of the question asking about indigenous identity.

Table 1. Racial amalgamation of racial self identification of students in demographic survey to the standard shown in Appendix part A

Original Declaration of Race	**Changed To**
Kyrgyz,Turkish,Israeli,Jewish,Kazakh,Persian,Persian/Baloch	Middle Eastern/Central Asian
Eastern Slavic,Russian,Slavic,Ukrainian,Caucasian	White
Aboriginal,First Nations, Métis	Indigenous
Japanese,Chinese,Koean,Taiwanese	East Asian
Latin American	Latino
Vietnamese,Filipino	Southeast Asian
Indian,Punjabi	South Asian
Mauritian,Nigerian	Black
Canadian	Prefer not to say

Statistical data package was R-Studio. Student assignment scores were calculated according to a standardized learning gain as follows:

Learning Gain = **Final Submission - Initial Submission / 1 - Initial Submission**

The standardized learning gain for each of the three sets was then assigned a label based on its position in overall assignments out of 8 in the semester, LS2, LS3, and LS6. Students who received a worse score on the final submission than the initial submission were rounded up to 0 for learning gains. Of the total of 787 students, LS2 had 464 counts of 0, LS3 had 381 counts, and LS6 had 322 counts TABLE. I used Kruskal-Wallis tests to determine differences in learning gains between demographic groupings. To look at the overall comparison in learning gains throughout the semester, a t-test was conducted to sample the difference between the pairings of

LS2-LS3, LS3-LS6, and LS2-LS6. A p-value of < 0.05 was used to determine significance in all tests.

Results

Data was obtained from the fall 2024 Biology 1020 Lab component of the course. The total number of students in the sample obtained was 1193. Interestingly enough, the population studied had a large number of individuals declaring She/Her pronouns, with 523/787 respondents affirming. The largest majority of students in the population were White (311), not LGBTQIA2S+ (654), Second generation Students (637), First year students (588), not International students (699), She/Her pronouns (523), English was a first language (527) and were not Indigenous (714) (Table 2).

Student Demographics:

Table 2: Student Group and Subgroup Survey Results At the University of Manitoba Fall 2024.

Demographics	Count	%		Demographics	Count	%
Race				**LGBTQIA2S+**		
Black	65	8.26		Yes	96	12.20
East Asian	36	4.57		No	654	83.10
Indigenous	20	2.54		Prefer not to say	37	4.70
Latino	11	1.40				
South Asian	113	14.36		**First-generation student**		
Southeast Asian	96	12.20		Yes	133	16.90
White	311	39.52		No	637	80.94
Middle East/West Asian	31	3.94		Prefer not to say	17	2.16
Prefer not to say	37	4.70				
Mixed Race	67	8.51		**Undergraduate Year**		
				1st	588	74.71

International Student				2nd	122	15.50
Yes	84	10.67		3rd	48	6.10
No	699	88.82		4th	15	1.91
Prefer not to say	4	0.51		Prefer not to say	14	1.78
Pronouns				**Indigenous Identity**		
He/Him	238	30.24		Yes	59	7.50
She/Her	523	66.45		No	714	90.72
They/Them	10	1.27		Prefer not to say	14	1.78
Prefer not to say	16	2.03				
English Is a 2nd Language						
yes	250	31.77				
no	527	66.96				
Prefer not to say	10	1.27				

Standardized Learning Gains

In the tests using the Kruskal-Wallis test (Table 4, Table 5, Table 6), there were no significant differences in learning gains among groups. In the student population, there was no significant difference between groups of students or within those groups when it came to the learning gains experienced by the populations studied. Though results were not significant, I did average out the learning gains found in all three assignments to see what the overall average increase in mean learning scores was for each subgroup within the larger groups . Of note, the highest scores that were more than .1 (1%) higher than other scores over the entire semester were; Pronouns-They/Them (mean=.374 n=10), 1stGenStudent-Yes (mean=.340 n=133), English2nd-Yes (mean=.335 n=250), Race-Indigenous (mean=.358 n=20), and Student year-2nd year (mean=.324 n=122)(Table 7).

In the next test I decided to look at overall standardized learning gains in the population and found that among the entire student population, learning gains were found to to be significant from pairings of LS2-LS3, LS3-LS6, and LS2-LS6 (Table 3), and the mean learning gains increase over the course of the semester (Figure 1)L2S: Mean-.20, Min-0, Q1-0, Median-0, Q3-0.42, Max-1. L3S: Mean .32, Min-0, Q1-0, Median-.008, Q3-0.67, Max-1. L6S: Mean .42, Min-0, Q1-0, Median-0.35, Q3-0.86, Max-1).

Figure 1: A boxplot showing standardized learning gains (Learning Gain = Final Submission - Initial Submission / 1 - Initial Submission) from L2S, L3S, and L6S with a max of 1.0. L2S, L3S, and L6S represent the assignment learning gain from the second, third, and sixth assignments in the semester that were submitted first to Peerceptiv and then second to UMLearn as a final submission.

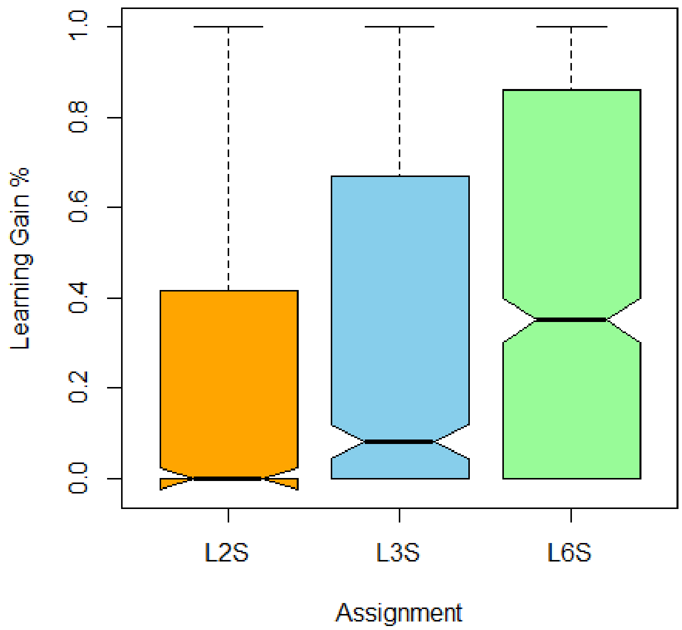

Table 3: Welch's Two-Sided T-Test Showing tests between pairings of L2S-L3S, L3S-L6S, and L2S-L6S, to test for significance in learning gains among all students.

Comparison	t-statistic	df	p value	CI lower	CI upper
L2S vs L3S	-7.274	1455.0	$p<.001$	-0.155	-0.089
L3S vs L6S	-5.282	1557.7	$p<.001$	-0.145	-0.067
L2S vs L6S	-12.729	1383.3	$p<.001$	-0.263	-0.193

Student Perceptions

Students overwhelmingly valued the peer-review process and answered affirmatively with either an agree or strongly agreeing that using peer review helped them learn science skills (80%), that getting reviews on their work was the most useful (87%), providing reviews on others' work was the most useful (76%) and that their understanding of concepts increased by using peer review (79%). Of those that felt negatively about the process, only 10% felt that peer review was a waste of time or that their work did not improve because of peer review 13%. Of the entire population, 83% would want peer review to be used in future assessments (Table 8).

Discussion

Overall, the learning gains improved over the course of the semester for all students, but mostly for those students identifying as Pronouns-They/Them (mean=.374 n=10), 1stGenStudent-Yes (mean=.340 n=133), English2nd-Yes (mean=.335 n=250), Race-Indigenous (mean=.358 n=20), and Student year-2nd year (mean=.324 n=122) (Table 7) .

Students did show overall increased learning gains over the course of the semester between these assignments, and the results were significant despite a large proportion doing worse than their initial submission. The process benefited all students approximately equally in the cohort, and students learned from prior mistakes in the semester or adapted to the format. This is in contrast to other studies that have shown a disproportionate learning gain in the student population where URM students have higher learning gains than NON-URM (Theobold et al. 2020). There were no significant differences between the groups of surveyed students when standardized for learning gains, demonstrating that all students benefited approximately in the same way (Table 4, Table 5, Table 6). In LS2, LS3, and LS6, students' performance remained relatively consistent with each other in spite of URM or NON-URM categorization. In the population, a high proportion of students did worse on the second submission after receiving a higher initial mark, suggesting factors in the process that need to be corrected for. Notably, if both the one submitting the assignment and the one reviewing take the process seriously we do see maximum gains in the learning gain (Biango-Daniels & Sarvary, 2021). When students had an initial submission to the Peerceptiv platform they received feedback from peers and then were given an opportunity to correct or improve the final submission. In some cases, students either did not correct for the final submission, or they got the same score as the initial submission. In either case the learning gain for those projects calculated was 0. Mean standardized learning scores in the population increased through the semester and show that initial vs second submissions after peer review increased the opportunity for students to learn from the process. Overall, there are positive trends in the effectiveness of peer review as a learning tool; however, methods for measuring the learning gains need improvement to ascertain which groups are benefiting the most.

Demographic Groups Show No Significant Differences

There were no significant results found within the demographic subgroupings of different students tested (Table 4, Table 5,Table 6). For example, the grouping of Pronouns contained within it four respondent categories; He/Him, She/Her, They/Them, and Prefer not to say (Table 2). Upon testing, no significant differences were found between any of the groupings or subgroupings. This is in contrast to other studies that found some groups benefited more than others significantly (Theobold et al., 2020; Strayhorn, 2009). Though this data was not significant, some groups had minor hi (Table 7), suggesting that the lack of statistical significance on this particular test means that some refinement must be done on the method for the collection and grading of the data. This could be due to a number of factors. In this cohort, the first round of submissions to Peerceptiv was benchmark graded and not as stringent in the grading process as the second set in the final submission. This had an effect on the measurable learning gains I could detect in the sample of 787 students. As stated prior, LS2 had 464 counts, LS3 had 381 counts, and LS6 had 322 counts of a 0 score or less for standardized learning gains.

It is of note that the number of leaning gains that equalled 0 decreased over the course of the semester. There are two possible explanations for this outcome. The first would be that students initially thought that their submission was good enough and simply turned it in again without reviewing. Students are shown to not be as critical of their own work as an expert would be (Biango-Daniels & Sarvary, 2021). Whether or not it is possible for further testing of this concept remains to be seen, as the labour and consistency required to grade over 1000 is higher than was available. In future studies, a smaller cohort might be selected at the beginning to have a pre-approved sample size that the same expert consistently grades, as this has been shown to act as a secondary control, allowing for a reference point to be established (Biango-Daniels & Sarvary, 2021). This would allow for a comparison of the expertly reviewed and graded vs. the ones done by peers and markers. Even if all of the peers are competent, none of them have expertise in education to the same level as the primary educator (Biango-Daniels & Sarvary, 2021; Andrews et al., 2011). The peer review that occurs from an instructor with a higher level of education giving feedback has been shown to be more accurate and viable for the quality provided (Biango-Daniels & Sarvary, 2021; Andrews et al., 2011). This has been shown to have an effect on the results of active learning in biological science as students overemphasise the knowledge that they have and miss things in the evaluation of their own and of peers' work (Andrews et al., 2011). That being said, the peer-review process still results in the occurrence of active learning even when an expert is not the one grading every single submission (Biango-Daniels & Sarvary, 2021.

All Students Benefited From Peer-Review

Our overall student population did show learning gains within assignments over the course of the semester (Table 3)(Figure 1). This supports the evidence that active learning by peer review increases understanding and retention in the population. The mean gains over time exhibit a steady increase in the learning gains of the student population regardless of demographic data (Theobold et al., 2020; Strayhorn, 2009).

As the semester goes on and they understand the system better, student learning may be skewed on low-risk assessments as they put forward only enough effort to get an equivalent reward (Salomon & Globerson, 1989). The initial submission then may suffer, and so a lower initial score is obtained resulting in the final submission having a higher learning gain. A higher value for assignments would likely have a better result in future iterations.

Student responses also have a high impact socially. Students fall into a range of categories from highly motivated and engaged to disengaged and less motivated (Salomon & Globerson, 1989). Socially, if a highly motivated student sees that they are not getting good feedback from peers, they may decide not to engage with the active learning activity as much in later tests. Students who do not feel that peers are engaging with the process of peer-review can become unmotivated by the lack of commitment to the process by peers and so decrease the quality of work to match the level of effort of peers (Salomon & Globerson, 1989). Lack of motivation may be remedied by increasing the rubric value for giving valuable feedback and giving instructions for what students should be looking for when they review as they give feedback. For those who do engage in the review process, either by giving or receiving feedback, learning gains are still expected to be similar to when expert feedback is given on the topic (Biango-Daniels & Sarvary, 2021). Our particular group of students responded very positively to the experience (Table 8).

Future Survey Considerations

In this data cohort, it was assumed that students who obtained a 0 did not engage with the process of peer-to-peer learning, as they did not turn in any assignment at all for grading or submitted the wrong document so those students were deleted from the data. For demographic choices, it was found that racial data had a smaller effect on our study than expected. Factors of race have wrapped up in them factors that influence student success, but the geographic or cultural origin of a person is not an accurate enough indicator in all cases for statistical study. Future surveys should likely ask for information from students about where they went to high school, financial income, if they work, or other indicators of coming from a less privileged situation that pulls from a student's mental resources and causes stress inhibiting learning, access to resources, and time allocation for education (Ong et al., 2019).In a study done on the psychological functioning of those with low income, those who had high amounts of debt every month experienced a bandwidth loss and a mental tax on decision making and capacity(Ong et al., 2019). The closest indicator found in this study for this type of information is likely found in the number of first-generation students in comparison to the overall population of 17%. Though this is not a direct indicator of income, students who are first-generation students may have less familial support than those who have family members who have higher learning degrees and consequently fewer financial resources.

In conclusion, the importance of studying how peer review affects student performance cannot be understated. Critical thinking and collaboration in our world an important social and intellectual skill that is a cornerstone of a functional society. Students benefit from the experience and it engages them on more than just the memorization of facts. Placing the responsibility for education in the hands of the community rather than only on the shoulders of a few. Investigating this gives students a varied and interactive experience that touches on themes they will experience later in both careers and the educational journey.

References

Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active Learning Not Associated with Student Learning in a Random Sample of College Biology Courses. *CBE—Life Sciences Education*, *10*(4), 394–405. https://doi.org/10.1187/cbe.11-07-0061

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing Diversity in Undergraduate Science: Self-Efficacy Drives Performance Gains with Active Learning. *CBE—Life Sciences Education*, *16*(4), ar56. https://doi.org/10.1187/cbe.16-12-0344

Bandura, A. (1977). *Social Learning Theory*. Prentice Hall.

Biango-Daniels, M., & Sarvary, M. (2021). A challenge in teaching scientific communication: Academic experience does not improve undergraduates' ability to assess their or their peers' writing. *Assessment & Evaluation in Higher Education*, *46*(5), 809–820. https://doi.org/10.1080/02602938.2020.1812512

Bonwell, C. C., & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports*. ERIC Clearinghouse on Higher Education, The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183 ($17. https://eric.ed.gov/?id=ED336049

Debets, Unpublished (2024)

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, *116*(39), 19251–19257. https://doi.org/10.1073/pnas.1821936116

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science (New York, N.Y.)*, *332*(6031), 862–864. https://doi.org/10.1126/science.1201783

Estrada, M., Burnett, M., Campbell, A. G., Campbell, P. B., Denetclaw, W. F., Gutiérrez, C. G., Hurtado, S., John, G. H., Matsui, J., McGee, R., Okpodu, C. M., Robinson, T. J., Summers, M. F., Werner-Washburne, M., & Zavala, M. (2016). Improving Underrepresented Minority Student Persistence in STEM. *CBE Life Sciences Education*, *15*(3), es5. https://doi.org/10.1187/cbe.16-01-0038

Fraser, J. M., Timan, A. L., Miller, K., Dowd, J. E., Tucker, L., & Mazur, E. (2014). Teaching and physics education research: Bridging the gap. *Reports on Progress in Physics. Physical Society (Great Britain)*, *77*(3), 032401. https://doi.org/10.1088/0034-4885/77/3/032401

Freeman, S., O'Connor, E., Parks, J. W., Cunningham, M., Hurley, D., Haak, D., Dirks, C., & Wenderoth, M. P. (2007). Prescribed Active Learning Increases Performance in Introductory Biology. *CBE—Life Sciences Education*, *6*(2), 132–139. https://doi.org/10.1187/cbe.06-09-0194

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74. https://doi.org/10.1119/1.18809

Helou, M. M., & Newsome, L. K. (2018). Application of Lev Vygotsky's Sociocultural Approach to Foster Students' Understanding and Learning Performance. *Journal of Education and Culture Studies*, *2*(4), 347. https://doi.org/10.22158/jecs.v2n4p347

Huisman, B., Saab, N., van den Broek, P., & van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: A Meta-Analysis. *Assessment & Evaluation in Higher Education*, *44*(6), 863–880. https://doi.org/10.1080/02602938.2018.1545896

Illeris, K. (2008). *Contemporary Theories of Learning: Learning Theorists ... in Their Own Words*. Routledge. http://ebookcentral.proquest.com/lib/umanitoba/detail.action?docID=425525

Maddox, G. B., Balota, D. A., Coane, J. H., & Duchek, J. M. (2011). The role of forgetting rate in producing a benefit of expanded over equal spaced retrieval in young and older adults. *Psychology and Aging*, *26*(3), 661–670. https://doi.org/10.1037/a0022942

Meyers, E. M., Qi, X.-L., & Constantinidis, C. (2012). Incorporation of new information into prefrontal cortical activity after learning working memory tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(12), 4651–4656.

National Academies of Sciences, Engineering, and Medicine, Barriers and Opportunities for 2-Year and 4-Year STEM Degrees: Systemic Change to Support Students' Diverse Pathways (The National Academies Press, Washington, DC, 2016).

Ong, Q., Theseira, W., & Ng, I. Y. H. (2019). Reducing debt improves psychological functioning and changes decision-making in the poor. *Proceedings of the National Academy of Sciences*, *116*(15), 7244–7249. https://doi.org/10.1073/pnas.1810901116

Orsmond, P., Merry, S., & Sheffield, D. (2006). A quantitative and qualitative study of changes in the use of learning outcomes and distractions by students and tutors during a biology poster assessment. *Studies in Educational Evaluation*, *32*(3), 262–287. https://doi.org/10.1016/j.stueduc.2006.08.005

Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research*, *13*(1), 89–99. https://doi.org/10.1016/0883-0355(89)90018-9

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why Peer Discussion Improves Student Performance on In-Class Concept Questions. *Science*, *323*(5910), 122–124.

Strayhorn, T. (2009). *Measuring The Educational Benefits Of Diversity In Stem Education: A Multi Institutional Survey Analysis Of Women And Underrepresented Minorities*. 14.870.1-14.870.9. https://peer.asee.org/measuring-the-educational-benefits-of-diversity-in-stem-education-a-multi-institutional-survey-analysis-of-women-and-underrepresented-minorities

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., … Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Topping, K. J. (2009). Peer Assessment. *Theory Into Practice*, *48*(1), 20–27. https://doi.org/10.1080/00405840802577569

Tables and Figures

Table 4: Kruskal-Wallis test for Assignment L2S testing for significance of learning gains

Assignment L2S

GroupingVariable	Statistic	DF	P-Value
Pronouns	0.121	3	0.9892
International Student	1.119	2	0.5716
English2nd	1.007	2	0.6045
Race	8.02	9	0.5321
Rainbow	1.644	2	0.4395
1stGenStudent	3.24	2	0.1979
Indigenous	0.436	3	0.9328

Table 5: Kruskal-Wallis test for Assignment L3S testing for significance of learning gains.

Assignment L3S

GroupingVariable	Statistic	DF	P-Value
Pronouns	4.359	3	0.2253
International Student	3.726	2	0.1552
English2nd	0.905	2	0.6359
Race	8.882	9	0.4482
Rainbow	2.535	2	0.2815
1stGenStudent	2.414	2	0.2992
Indigenous	1.511	3	0.6797

Table 6: Kruskal-Wallis test for Assignment L6S testing for significance of learning gains.

Assignment L6S

GroupingVariable	**Statistic**	**DF**	**P-Value**
Pronouns	5.197	3	0.1579
International Student	2.942	2	0.2297
English2nd	2.879	2	0.2371
Race	3.061	9	0.9618
Rainbow	0.443	2	0.8014
1stGenStudent	1.054	2	0.5905
Indigenous	1.423	3	0.7001

Table 7: Student Grouping and subgrouping means of all three assignments

	Mean ALL	**std dev**	**n**			**Mean ALL**	**std dev**	**n**
Pronouns					***Rainbow***			
He/his	0.311	0.209	238		No	0.313	0.224	654
She/her	0.311	0.227	523		Yes	0.307	0.219	96

They/Them	0.374	0.242	10					
					1stGenStudent			
InternationalStudent					No	0.305	0.221	637
No	0.314	0.225	699		Yes	0.340	0.225	133
Yes	0.307	0.205	84					
					English2nd			
Race					No	0.302	0.222	527
Black;	0.317	0.209	65		Yes	0.335	0.223	250
East Asian;	0.334	0.182	36					
Indigenous;	0.358	0.247	20		***StudentYear***			
Latino;	0.279	0.137	11		1st year	0.309	0.224	588
Middle Eastern/West Asian	0.301	0.229	31		2nd year	0.324	0.224	122
Mixed Race;	0.270	0.227	67		3rd year	0.306	0.201	48
South Asian;	0.325	0.232	113		4th year or more	0.302	0.178	15
Southeast Asian;	0.297	0.231	96					
White;	0.318	0.223	310		***Indigenous***			
					No	0.313	0.222	714

					Yes	0.310	0.235	59

Table 8: Student Opinions on Peer Review Effectiveness

Question	Response	n	%
Using Peer Review Helped Me Learn Science Skills			
	Strongly Agree	212	26.9
	Agree	419	53.2
	Diagree	85	10.8
	Strongly Disagree	24	3.0
	N/A	47	6.0
Getting Reviews On My Work Was The Most Useful			
	Strongly Agree	361	45.9
	Agree	320	40.7
	Diagree	66	8.4
	Strongly Disagree	14	1.8
	N/A	26	3.3
Providing Reviews On Others Work Was The Most Useful			
	Strongly Agree	176	22.4
	Agree	419	53.2
	Diagree	122	15.5
	Strongly Disagree	17	2.2
	N/A	53	6.7
My Understanding Of Concepts Increased By Using Peer Review			
	Strongly Agree	257	32.7

	Agree	364	46.3
	Diagree	96	12.2
	Strongly Disagree	21	2.7
	N/A	49	6.2
I Felt That Peer Review Was A Waste Of Time			
	Strongly Agree	24	3.0
	Agree	51	6.5
	Diagree	390	49.6
	Strongly Disagree	279	35.5
	N/A	43	5.5
My Work Did Not Improve Because Of Peer Review			
	Strongly Agree	25	3.2
	Agree	77	9.8
	Diagree	416	52.9
	Strongly Disagree	222	28.2
	N/A	47	6.0
I Would Want Peer Review To Be Used In Future Assessments			
	Strongly Agree	330	41.9
	Agree	323	41.0
	Diagree	49	6.2
	Strongly Disagree	24	3.0
	N/A	61	7.8

Appendix.

Appendix Part A.

Questions presented to students for the survey.

Survey
What are your preferred pronouns?
He/his
She/her
They/Them
Prefer not to say

Are you an International Student?
No
Yes
Prefer not to say

Do you identify as: (select more than one if applicable)
Black;
East Asian;
Indigenous;
Latino;
Middle Eastern/West Asian
Mixed Race;
South Asian;
Southeast Asian;
White;
Other;
Prefer not to say;

Do you consider yourself a part of the LGBTQIA2S+ Community?
No
Yes
Prefer not to say

Are you a first-generation university student? (the first person in your family to attend post-secondary education)
No
Yes
Prefer not to say

Is English your second language?
No
Yes
Prefer not to say

What year of this undergraduate degree are you in?
1st year
2nd year
3rd year
4th year or more
Prefer not to say

Do you Identify as Indigenous?
No
Yes
Prefer not to say

Perceptions survey

Using Peer review helped me learn science skills
Getting reviews on my work was the most useful
Providing reviews on others' work was the most useful
My understanding of concepts increased by using peer review
I felt that peer review was a waste of time
I would want peer review to be used in future assessments
My work did not improve because of peer review

All questions were answered as: Strongly Agree, Agree, Disagree, Strongly Disagree.

Appendix B. Assignment Rubric

Iechanics (/7 Marks)	
Descriptive Caption	Mark - Should be: a) in the form of a statement, b) mention the subject(s), c) appropriate ıriables including a key to their representation (i.e., colour or pattern), d) include relevant ɛtails about the experiment essential for understanding the take home message .75 Mark - The caption is present but missing one of the above criteria. .5 Mark - The caption is present but missing two of the above criteria. .25 Mark - The caption is present but missing three of the above criteria. Marks – Caption is not clearly identified or communicated
X Axis	Mark – Includes descriptive axis label and proper units .5 Mark – Is missing either axis label or proper units Mark – Missing both axis label and units
Y Axis	Mark – Includes descriptive axis label and proper units .5 Mark – Is missing either axis label or proper units Mark – Missing both axis label and units
Scale	Mark - Intervals and range should be appropriate for the data displayed. .5 Mark - The scale is not appropriate for one of the axes. Mark -The scale is not appropriate for any axis.
Legend	Mark - Should be appropriate and descriptive for the data displayed and include descriptions ' different colours/patterns .5 Mark - The legend is not descriptive or includes errors Mark – The legend is not present, and is needed
Error bars	Mark – Error bars are needed with the data, and are clearly presented or Error bars are not ɛeded with the dataset and are not included in the graph .5 Mark – The errors bars are not accurate

	Mark – There are no error bars & they are needed; there are error bars when there should not e
raph Choice (/ 4 Marks)	
ase of nderstanding	Marks - The graph is constructed in a way that is: a) clear to identify trends and b) easy to note e take home message Mark – The graph is constructed in a way that data trends are difficult to observe or it is fficult to understand the take home message. Marks – The graph is ineffective at communicating trends and the take home message.
raph Type	Marks - The graph has: a) data plotted to take up sufficient room in the Cartesian plane, b) gible fonts, c) clear and legible X- and Y- axis lines, d) data displayed in the appropriate ımber of bars/lines, and e) no junk elements such as distracting colours, patterns, dark grid nes, and is appropriate for the dataset. .5 Marks – The graph is missing one of the points above. Mark – The graph is missing two of the points above. .5 Mark – The graph is missing three of the points above. Marks – The graph has too many flaws that interfere with the interpretation of the data

Summary (/10 Marks)	
Length is appropriate	3 marks – Good length (1 paragraph and less than half a page) 2 marks – The summary is too long (could be written more concise) 1 marks – The summary is too vague, and needs more details provided. 0 mark – There is no summary included
Writing Ability	3 marks – There were no grammatical errors in the document. 2 mark – There were a few grammatical errors in the document. 1 mark – There were 4 – 5 minor grammatical errors in the document. 0 mark – There were lots of grammatical errors in the document.

Accuracy	3 marks – There summary is an accurate match to the data provided in the graph. 2 mark – There were a few statements that do not match the data provided in the graph. 1 mark – There were many statements that do not match the data provided in the graph. 0 mark – The summary does not describe or match the data provided in the graph.
Correctness	1 mark – The summary contained no incorrect statements. 0.5 marks – There were 1 – 3 incorrect statements. 0 marks – The summary was not correct.