A Learning Analytics-informed Activity to Improve Student Performance in a First Year Physiology Course

Mark T. Williams[a], Lesley J Lluka[a] and Prasad Chunduri[a]

Corresponding author: Prasad Chunduri (p.chunduri@uq.edu.au)
[a]School of Biomedical Sciences, The University of Queensland, Brisbane QLD 4072, Australia

Keywords: short answer question (SAQ), learning analytics intervention, evaluative judgement

Abstract

Learning Analytics (LA) can be employed to identify course-specific factors that hinder student course (outcome) performance, which can be subsequently rectified using targeted interventions. Supplementing interventions with predictive modelling also permits the identification of students who are at-risk of failing the course and encourages their participation. LA findings suggested that a targeted intervention for our course should focus on improving student short answer question (SAQ) performance, which we attempted to achieve by improving their understanding of features pertaining to various SAQ answer standards and how to achieve them using examples of varying scores. Every student was invited to the intervention via a course-wide announcement through the course learning management system. At-risk students identified using predictive models were given an additional invitation in the form of a personalised email. Results suggest that intervention improved student understanding of SAQ performance criteria. The intervention also enhanced student end-of-semester SAQ performance by 12% and 11% for at-risk and no-risk students respectively. Course failure rate was also lower by 26% and 9% among at-risk and no-risk intervention participants. Student perception of the intervention was also positive where an overwhelming majority of participants (96%) found the interventional activity to be useful for their learning and exam preparations.

Introduction

Factors affecting students' academic performance at universities are multifaceted (Tinto, 2006) and are often specific to the unit/course that they are enrolled in (Gašević, Dawson, Rogers, & Gasevic, 2016; Vermunt, 2005). Evidence suggests that embedding interventions targeted towards course-specific issues would be highly effective in improving student outcomes in that particular course and overall academic performance (Gašević et al., 2016; Rathner, Hughes, & Schuijers, 2013; Tangalakis, Best, & Hryciw, 2017; Vermunt, 2005). A targeted intervention would also ensure that benefits are maximised with minimal investment of additional student time and effort – assets that are limited in large, content heavy courses like ours. These issues can be identified using Learning Analytics (LA), a multidisciplinary field that enables the identification of significant trends in student-generated data through deductive and inductive means (e.g., Gašević et al., 2016). There is growing evidence supporting the validity of LA-based interventions at a course-level to improve student outcomes (Megaw & Zimanyi, 2019; Toetenel & Rienties, 2016; van Leeuwen, Janssen, Erkens, & Brekelmans, 2015). Results from LA are also useful to construct predictive models (Gašević et al., 2016) that allow instructors to identify students at-risk of low course performance and outcome, and subsequently encourage their participation in such targeted interventions.

Generally, assessment tasks are a suitable target for interventions due to their influence on not only the students' course outcomes, but also on their learning process (Biggs, 2003; Scouller, 1998). Optimising the learning process is paramount to improving student course trajectory, particularly for low performing students as they were shown to adopt learning practices that foster poor conceptual understanding and course performance (Hazel, Prosser, & Trigwell, 2002). Insights from course coordinators, further guided by LA findings and literature, can be used to identify assessment tasks that are most suited for the intervention, particularly high stakes tasks that students find challenging. In physiology courses, where development of complex conceptual understanding is a fundamental component of most course learning objectives, this would most apply to the short answer question (SAQ) assessment. SAQs are inherently difficult for students, particularly for first year university students, due to their intrinsic task requirements (Carnegie, 2015; Rozenblit & Keil, 2002; Sefton, 1998), such as:

(1)Integration of concepts across course modules and biological systems to ensure that concepts are described at the required depth
(2)Inclusion of concepts that are relevant to the question
(3)Ensuring that ideas are organised into coherent sentences

This difficulty is compounded further in courses where SAQs are only presented to students in the end of semester (EOS) exam, as students are not given an opportunity to practice answering SAQs or given instructor feedback. This prevents students from gaining insights into task expectations and salient features of high-scoring answers, and thus hindering their capacity to adequately prepare for the task (Sadler, 1989; Scouller, 1998).

Examples are effective in enhancing student understanding of task expectations and quality indicators (Hendry, Armstrong, & Bromberger, 2012; Hendry, Bromberger, & Armstrong, 2011; Rust, Price, & O'Donovan, 2003; Wimshurst & Manning, 2013; Yucel, Bird, Young, & Blanksby, 2014), and thus would serve as a suitable basis for our intervention. Engaging students with these examples, often through iterative instructor-led discussions, have been shown to improve student evaluative judgement – their capacity to evaluate their own work and that of others, and more importantly, their future performance in related tasks (Carless & Chan, 2016; Carless, Ho Chan, To, Lo, & Barrett, 2018; Tai, Ajjawi, Boud, Dawson, & Panadero, 2018; To & Carless, 2016). We have, in the past, implemented an intervention that led to improvements in student SAQ performance (Williams, Lluka, Meyer, & Chunduri, 2019). However, feedback provided in the previous activity was too intricate to represent conventional answers to SAQs. This is because the examples provided were essay-length that integrated concepts from every module in the course, while high-scoring SAQ responses in our course only require integration of two modules at most. Furthermore, the previous intervention was not supplemented with LA-based predictive model(s), which prevents us from effectively targeting students who would benefit most from the activity, specifically students who are at-risk of failing the course. Lastly, the activity was conducted towards the end of the teaching semester, specifically in the last week of classes. Thus, a targeted activity that is implemented relatively early in the semester was necessary to provide students with a clearer example of how SAQs would be structured in the EOS exam, and how best to answer them.

Taking these into account, the current study attempted to implement an intervention to improve student performance in SAQs, a clear drawback in students' course performance as evidenced by LA results and course coordinators' insights (Supplementary Material: Appendix A). This was achieved by providing students with annotated examples of SAQ answers of varying standards to articulate task expectations and features of high-scoring answers. Standards were disseminated in this manner as certain key features of high-scoring answers, such as coherence and the extent to which concepts are integrated, can be effectively articulated only through examples (Orsmond, Merry, & Reiling, 2002; Sadler, 1989). Feedback was also framed in this manner to refine student capacity to discriminate between different performance levels and to perform the task competently (Kitsantas & Zimmerman, 2006; Orsmond et al., 2002). Unlike previous studies, our intervention did not include instructor-led student discussions of the examples (e.g., Rust et al., 2003; To & Carless, 2016; Yucel et al., 2014) but instead, students were given a greater number of examples (6 per question, rather than the usual 2-3). This was done to further reduce the additional investment of time and effort by students in our course, as such iterative discussions are time consuming, which in turn may discourage student participation in the activity. The current study also attempted to utilise predictive models to identify at-risk students and subsequently encourage their participation in the intervention.

Methods

Course context

Ethical clearance for the current study was obtained from The University of Queensland (UQ) Human Behavioural and Social Sciences Ethical Review Committee. The current study was conducted in 2016 and 2017, in the first semester offering of a large first year physiology course at UQ, Australia. There were 611 and 642 students enrolled in the course in 2016 and 2017 respectively. This course covered and emphasised the integration of key biological processes in various physiological systems that are critical for the functioning of complex organisms, particularly humans. Concepts were disseminated through content lectures and reinforced through active learning workshops, peer-assisted study sessions, and laboratory practical sessions throughout the semester.

Assessment in the course includes practical reports, quizzes, and an EOS exam, which consists of multiple-choice questions (MCQs) and SAQs. The EOS exam, particularly some of the SAQs, evaluates the extent to which students develop a holistic and integrative understanding of the course content. Typically, the EOS exam includes three SAQs, each with multiple sub-questions which are contextualised using a scenario description. SAQs are prepared by experts responsible for teaching the content. The assessment weightings are summarised in Table 1 with the exception of the Practical Core Competencies component, which assesses student mastery of four laboratory techniques. Students are given multiple attempts to demonstrate that they are able to perform these techniques satisfactorily and unassisted, and passing is a hurdle requirement to pass the course.

Table 1. Overview of course assessment components and the underlying tasks.

Component	Task name/theme	Abbreviation	Task description	Grading
Practical	Toad Anatomy	Prac 1	Laboratory report	5% each and each assessed using a set of scoring criteria
	Osmosis	Prac 2		
	Action Potential	Prac 3		
	Skeletal Muscle	Prac 4		
	Integration	Prac 5	Concept map	
Knowledge	Quiz 1	Quiz 1	Questions based on concepts taught in course	5% each
	Quiz 2	Quiz 2		
	Quiz 3	Quiz 3		

	EOS exam: MCQs and SAQs	EOS		60%:
				SAQs = 20%;
				MCQs = 40%

Developing an intervention targeted towards improving student course performance

Performance of students who passed the course in 2016 was compared to those who failed across all course assessment tasks. Descriptive statistics, discourse with course coordinators (instructors who design and manage the course), and literature, were all used to guide subsequent comparisons of student performance between and within certain assessment tasks (see Appendix A(I)). Results from the analysis and literature presented in the Introduction, suggests that the intervention in the 2017 course iteration should focus on improving student SAQ performance as students, especially those who failed were found to perform the worst in this Knowledge component relative to other assessment tasks ($p<0.001$ for all comparisons, Supplementary Material: Appendix A(I)).

Construction of predictive models

Briefly, predictive logistic regression models were constructed using the methodologies described in Harrell (2015) and Fox and Weisberg (2011). Estimated coefficients were tested for significance using the Wald test (Fox & Weisberg, 2011; Harrell, 2015). Model predictive capacity was summarised using the Brier score and the Area Under the Receiver Operating Curve (AUCROC), estimated using the bootstrap resampling method as recommended by Harrell (2015). The following variables were included in the models:

- All assessment tasks that have been completed by students at the point when the model is implemented during the semester
    - Prac 1, Prac 2, Quiz 1, and Quiz 2
- Pre-university data

- ATAR – A rank assigned to each student based on their performance in Australian high schools (1-99, worst to best)
- Chem HS – Variable denoting whether students studied at least one high school chemistry subject (coded 1 for yes)
- Nationality – domestic Australian students (coded as 1) or international students.

Two models were constructed because a small proportion of students were lacking pre-university data (12%). Pre-university data was not excluded from the model building process as model predictive capacity improved when said data was included (see Supplementary Material: Appendix A(II)) and models can only predict course outcomes of students who have data pertaining to every variable in a model. All models were considered to have excellent capacity to discriminate between the two outcomes due to AUCROC being within the 0.8-0.9 range (Hosmer & Lemeshow, 2000). All models fulfilled regression assumptions.

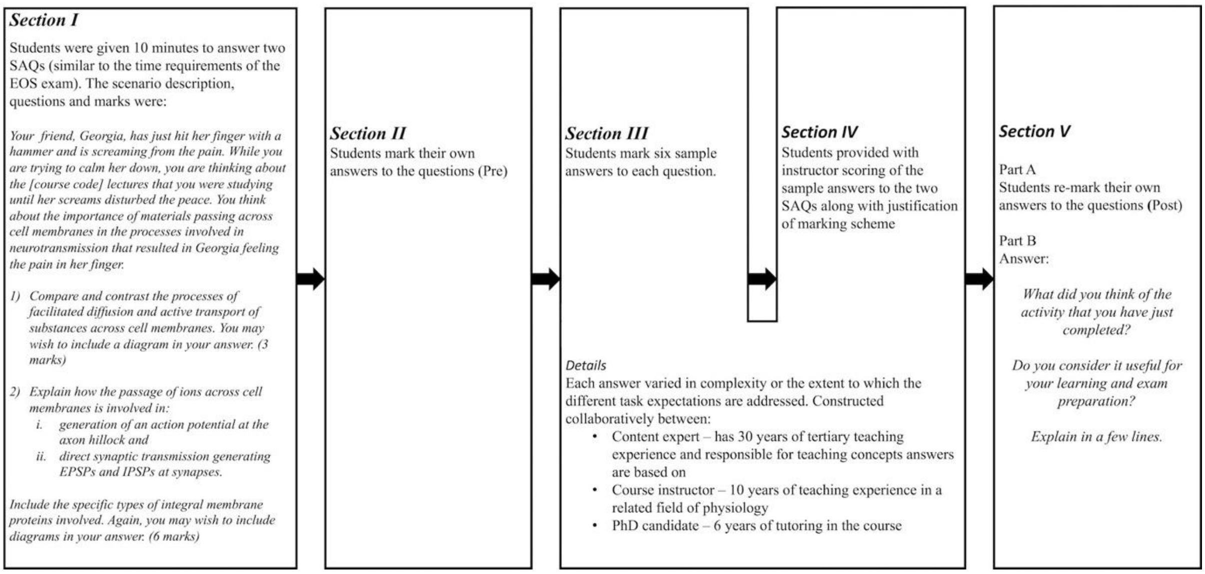
The predictive models (Appendix A(II)) were used to identify students who were at-risk of failing the course in 2017. These students were then encouraged to participate in the intervention via a personal email, which was in addition to the notification received by every enrolled student through an announcement via the university's online learning management system (Supplementary Material: Appendix B).

Interventional SAQ activity

The interventional activity was conducted during one of the lectures in week 10 of the conventional 13-week teaching semester. The activity consisted of five different sections, I to V (see Figure 1 for an overview, Supplementary Material: Appendix C for instructor sample answers).

Section I served to expose students to their own capacity to answer SAQs and as a means to self-evaluate. The first SAQ that students answered in Section I, Q1 was based on concepts taught in the first module of the course, specifically the processes by which substances of different properties move across the cell membrane. The second SAQ, Q2 was more complex and required students to integrate information from Module 1 and 2 to explain how signals are propagated from one neuron to another. In Section II, students were tasked to self-evaluate their performance in the two questions to serve as a baseline measure of student self-evaluation accuracy and a proxy of the internal scoring schemas that they utilise to judge the quality of SAQ answers. Sections III and IV served to calibrate student scoring schemes to better match that of the instructor by improving student understanding of the task expectations and indicators of SAQ answer quality. The range of answers provided to students highlighted the different levels of competence and the expected standards to be met for each competency level. Each answer was assigned to a particular performance standard through a unique score, which was given to the students verbally when their scoring of these alternative answers was completed. Verbal instructor feedback also highlighted the importance of including only relevant concepts in the answer, issues with verbosity, use of diagrams to demonstrate understanding, the depth of knowledge required, and importance of conceptual integration in achieving high-scoring SAQ answers. Section V consists of Part A and B, where students were required to self-score their SAQ answers in Part A in the same manner as Section II, while Part B was used to obtain students' views on the effectiveness of the activity in assisting with their preparation for the EOS exam. Students were permitted to take home the sheet with the instructor-generated answers and scoring, for review at a later time.

Finally, performance in the SAQs of the EOS exam of activity participants was compared to non-participants. The SAQs in the EOS exam were not directly related to the concepts covered in this formative activity.



Section I

Students were given 10 minutes to answer two SAQs (similar to the time requirements of the EOS exam). The scenario description, questions and marks were:

Your friend, Georgia, has just hit her finger with a hammer and is screaming from the pain. While you are trying to calm her down, you are thinking about the [course code] lectures that you were studying until her screams disturbed the peace. You think about the importance of materials passing across cell membranes in the processes involved in neurotransmission that resulted in Georgia feeling the pain in her finger.

1) *Compare and contrast the processes of facilitated diffusion and active transport of substances across cell membranes. You may wish to include a diagram in your answer. (3 marks)*

2) *Explain how the passage of ions across cell membranes is involved in:*
   i. *generation of an action potential at the axon hillock and*
   ii. *direct synaptic transmission generating EPSPs and IPSPs at synapses.*

Include the specific types of integral membrane proteins involved. Again, you may wish to include diagrams in your answer. (6 marks)

Section II

Students mark their own answers to the questions (Pre)

Section III

Students mark six sample answers to each question.

Section IV

Students provided with instructor scoring of the sample answers to the two SAQs along with justification of marking scheme

Details

Each answer varied in complexity or the extent to which the different task expectations are addressed. Constructed collaboratively between:

- Content expert – has 30 years of tertiary teaching experience and responsible for teaching concepts answers are based on
- Course instructor – 10 years of teaching experience in a related field of physiology
- PhD candidate – 6 years of tutoring in the course

Section V

Part A

Students re-mark their own answers to the questions (Post)

Part B

Answer:

What did you think of the activity that you have just completed?

Do you consider it useful for your learning and exam preparation?

Explain in a few lines.

Figure 1: Formative SAQ exercise outline.

Analysis

Student data was de-identified prior to analysis. Students in the 2017 iteration were allocated into different groups according to the selection criteria listed in Figure 2. The analysis plan and the specific student groups used to address the various study aims are also presented in Figure 2. Students in the 2017 cohort who were predicted to fail and pass the course were grouped as at-risk and no-risk students respectively. The responses of students who completed every section of the formative SAQ activity were scored by the content expert. This served to represent students' actual performance in the interventional SAQs, and as a basis to assess the self-evaluation accuracy of students. As the accuracy metric was calculated by subtracting instructor score from the student score, a positive accuracy metric indicates that students are overestimating the complexity of their own answer, while a negative metric indicates that students are underestimating it. Data is presented as median with interquartile range (IQR) unless stated otherwise. Inflated Type I error generated from repeated testing on each dataset was corrected through multiple comparisons tests or by adjusting the p-value using the Bonferroni method: Adjusted (p-value) = 0.05 (original p-value) / number of hypothesis tested, when such multiple comparisons tests are unavailable. All statistical tests used either $p<0.05$ or an adjusted p-value as threshold for significance.

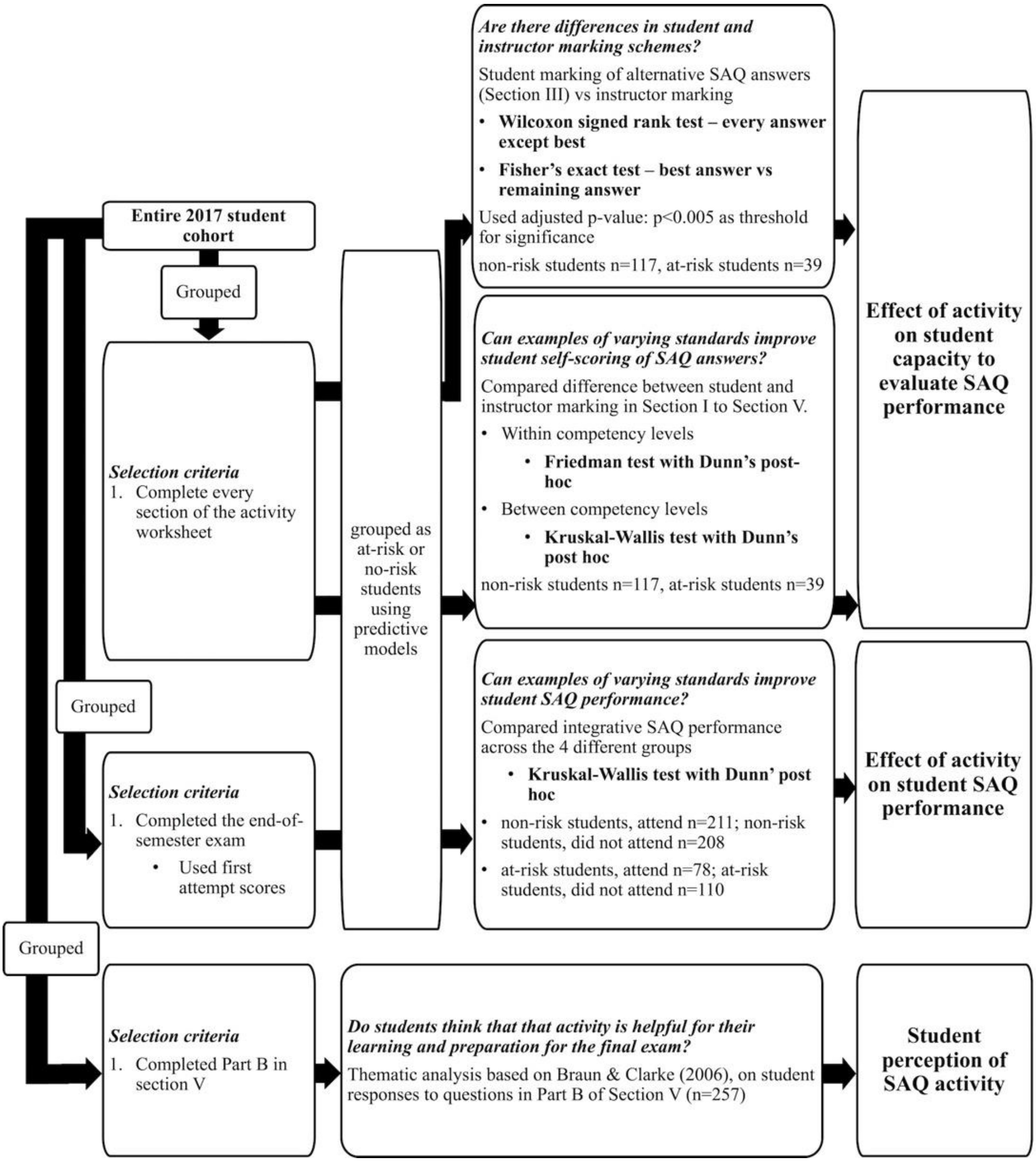

Entire 2017 student cohort
Grouped
Selection criteria
1. Complete every section of the activity worksheet
Grouped
Selection criteria
1. Completed the end-of-semester exam
• Used first attempt scores
Grouped
Selection criteria
1. Completed Part B in section V
grouped as at-risk or no-risk students using predictive models
Are there differences in student and instructor marking schemes?
Student marking of alternative SAQ answers (Section III) vs instructor marking
• Wilcoxon signed rank test – every answer except best
• Fisher's exact test – best answer vs remaining answer
Used adjusted p-value: p<0.005 as threshold for significance
non-risk students n=117, at-risk students n=39
Can examples of varying standards improve student self-scoring of SAQ answers?
Compared difference between student and instructor marking in Section I to Section V.
• Within competency levels
• Friedman test with Dunn's post-hoc
• Between competency levels
• Kruskal-Wallis test with Dunn's post hoc
non-risk students n=117, at-risk students n=39
Effect of activity on student capacity to evaluate SAQ performance
Can examples of varying standards improve student SAQ performance?
Compared integrative SAQ performance across the 4 different groups
• Kruskal-Wallis test with Dunn' post hoc
• non-risk students, attend n=211; non-risk students, did not attend n=208
• at-risk students, attend n=78; at-risk students, did not attend n=110
Effect of activity on student SAQ performance
Do students think that that activity is helpful for their learning and preparation for the final exam?
Thematic analysis based on Braun & Clarke (2006), on student responses to questions in Part B of Section V (n=257)
Student perception of SAQ activity

Figure 2: Selection criteria used to allocate students into different groups and analysis used to address the different study aims. The Wilcoxon signed rank test was not used to evaluate the accuracy of student scoring of the best answers for both questions in Section III as data did not fulfil a test assumption (i.e., it is not possible for student assigned scores to be distributed around the median score of 3 and 6 for Q1 and Q2 respectively).

Results

Feedback improved student self-scoring of their SAQ answers

Student self-scores for their answer to Q1 and Q2 in Section V was compared to Section II to determine if instructor feedback and examples improved student evaluation of their own SAQ performance. Analysis revealed that both at-risk and no-risk students were more accurate in their self-scoring for Q1 after being given feedback ($\text{PreQ1}_{\text{at-risk}}$: 0.5[IQR, 0 to 1], $\text{PostQ1}_{\text{at-risk}}$: 0[IQR, -0.5 to 0.5], p<0.01; $\text{PreQ1}_{\text{no-risk}}$: 1[IQR, 0.5 to 1]; $\text{PostQ1}_{\text{no-risk}}$: 0.5[IQR, -0.5 to 1], p<0.001; Figure 3). A similar but less pronounced improvement in accuracy was also observed for Q2, where feedback only managed to reduced student overestimation of their Q2 performance ($\text{PreQ2}_{\text{at-risk}}$: -0.5[IQR, -1.5 to 1.5], $\text{PostQ2}_{\text{at-risk}}$: -0.5[IQR, -1.5 to 0.5], p<0.05; $\text{PreQ2}_{\text{no-risk}}$: 0.5[IQR, -0.5 to 1.5]; $\text{PostQ2}_{\text{no-risk}}$: 0.5[IQR, -0.5 to 0.5], p<0.001; Figure 3).

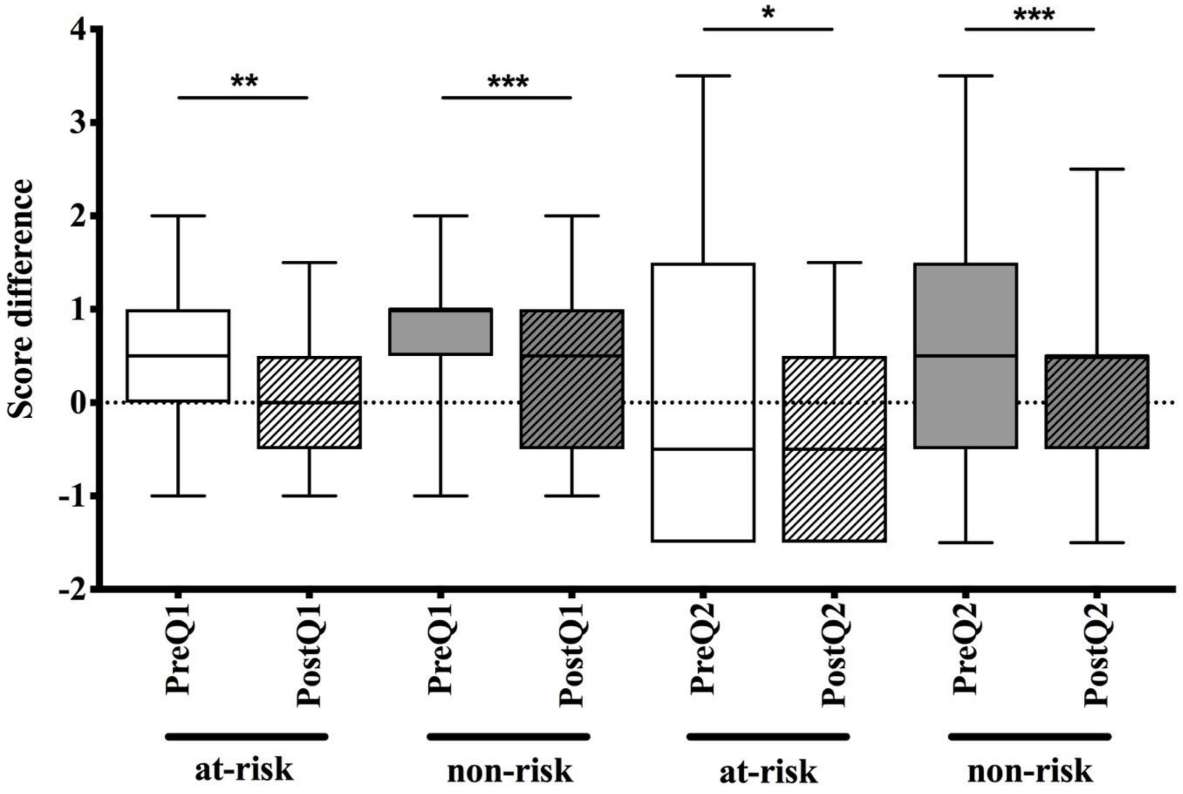


Figure 3: Changes in student scoring accuracy of their own answers to the interventional SAQs. Differences between the instructor assigned score and student self-assigned score to Q1 and Q2, prior to (Pre) and after (Post) exposure to example answers and instructor feedback are shown. Data is depicted as median with IQR. The Friedman test with Dunn's multiple comparisons test was used to determine if Pre-scores were significantly different to Post-scores for both the at-risk and no-risk students. The Kruskal-Wallis with Dunn's multiple comparisons test was used to compare marking score differences between no-risk and at-risk students. N(no-risk) = 117, N(at-risk) = 39. * p<0.05 **p<0.01 *** p<0.001

Distribution of student scores for most alternative answers deviated significantly from that of the instructor (Figure 4). For Q1, student score distribution for answers with score 0.5 and 1.5 appear to be significantly higher than that of the instructor for both at-risk and no-risk students ($\text{Score0.5}_{\text{at-risk}}$: 1.5[IQR, 1 to 2], $\text{Score0.5}_{\text{no-risk}}$: 1[IQR, 0.5 to 1.5], $\text{Score1.5}_{\text{at-risk}}$: 2.5[IQR, 2 to 2.5], $\text{Score1.5}_{\text{no-risk}}$: 2[IQR, 1 to 2.5]; $p<0.001$ for all comparisons; Figure 4A - 4D). Scores assigned by at-risk and no-risk students to model answer 1 had a median of 1[IQR, 1.5 to 0.5] ($p>0.005$, Figures 4A - 4D). The remaining answers for both student groups, except the answer with score 3, had a student score distribution that was significantly lower than the instructor-assigned score ($\text{Score2}_{\text{at-risk}}$: 1.5[IQR, 1 to 2], $\text{Score2}_{\text{no-risk}}$: 1.5[IQR, 1 to 2], $\text{Score2.5}_{\text{at-risk}}$: 2[1.5 to 2.5], $\text{Score2.5}_{\text{no-risk}}$: 2[IQR, 2 to 2.5]; $p<0.001$ for all comparisons; Figure 4A - 4D). Furthermore, less than 50% of both student groups were able to identify the correct score for the answers except for the best answer for Q1 ($\text{Score3}_{\text{at-risk}}$: 74%, $\text{Score3}_{\text{no-risk}}$: 81%; $p<0.001$ for all comparisons; Figures 4B and 4D).

The distribution of student scores of Q2 answers with score 1 and 2 were significantly higher than that of the instructor for both at-risk and no-risk students ($\text{Score1}_{\text{at-risk}}$: 3[IQR, 2 to 5], $\text{Score1}_{\text{no-risk}}$: 3[IQR, 2 to 4], $\text{Score2}_{\text{at-risk}}$: 4[IQR, 2 to 5], $\text{Score2}_{\text{no-risk}}$: 3[IQR, 2 to 4]; $p<0.001$ for all comparisons; Figure 4E – 4H). Score distribution of Q2 answers with score 3 was significantly lower than that of the instructor for both student groups ($\text{Score3}_{\text{at-risk}}$: 2[IQR, 1 to 3], $\text{Score3}_{\text{no-risk}}$: 1.5[IQR, 1 to 3]; $p<0.001$; Figures 4E – 4H). The number of students who correctly identified the score of each sample answer to Q2 was less than 50% for both at-risk and no-risk students (Figures 4F and 4H). Interestingly, 77% and 84% of the at-risk and no-risk group respectively assigned a lower score to the best answer for Q2 (Figures 4F and 4H).

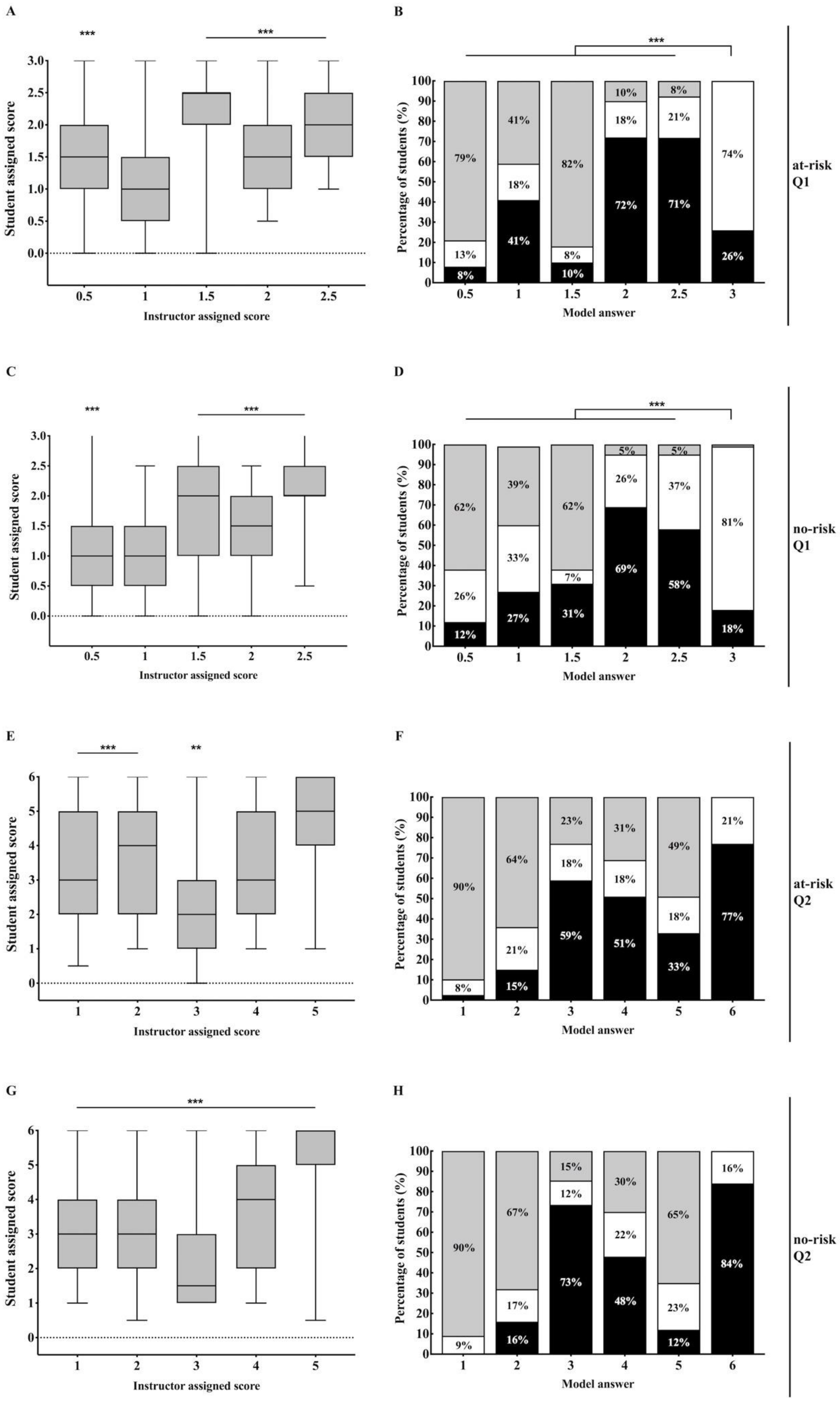


Figure 4: Student scoring of instructor sample answers. For each sample answer, the

proportion of students that assigned a score that is lower, match, and higher than the instructor for each answer are presented as black, white, and grey bars respectively in B, D, F, and G. % in bars represent the proportion of the total number of students that falls into each of the 3 scoring categories. Instructor assigned scores were used to represent complexity of sample answers with higher scores corresponding to greater complexity. The Wilcoxon signed rank test was used to determine if student scoring of each answer was significantly different to the instructor's. The proportion of students that correctly scored the best sample answer for both questions were compared with the remaining answers using Fisher's exact test. p-value required for significance was adjusted to p<0.005 using the Bonferroni method, to correct for increased Type 1 error due to repeated testing. N(at-risk) = 39, N(no-risk) = 117. ** p<0.005 *** p<0.001

Example answers of varying standards with feedback improved student SAQ performance

At-risk and no-risk students who attended the activity outperformed their non-attending counterparts in the SAQs of the EOS exam ($\text{Attend}_{\text{at-risk}}$: 39%[IQR, 27% to 50.5%], $\text{noAttend}_{\text{at-risk}}$: 27%[IQR, 15% to 40%]; p<0.05; $\text{Attend}_{\text{no-risk}}$: 62%[IQR, 47% to 75%], $\text{noAttend}_{\text{no-risk}}$: 50%[IQR, 37% to 65%]; p<0.001; Figure 5). However, no-risk students, whether they attended the activity or not, significantly outperformed at-risk students who attended the activity and those who did not (p<0.01 to p<0.001, Figure 5). Additionally, Fisher's exact test revealed that at-risk and no-risk students who attended the activity had a lower course failure rate than their non-attending counterparts (at-risk: 34% against 60%, p<0.001; no-risk: 6% against 17%, p<0.001).

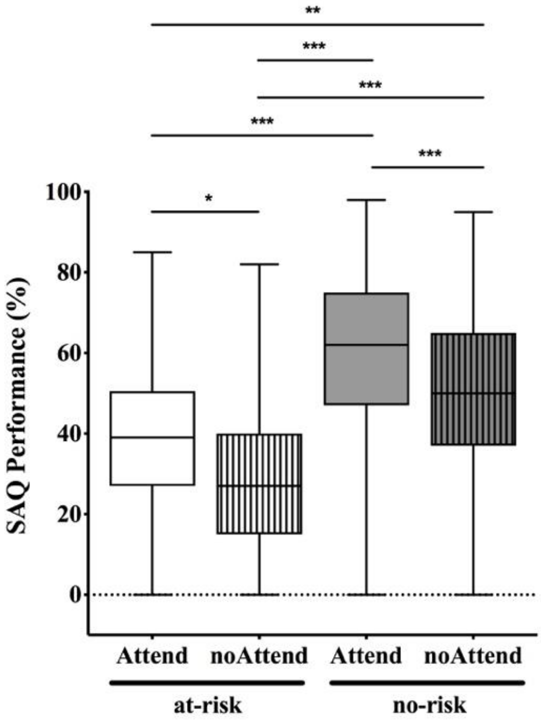


Figure 5: Student EOS exam SAQ performance. Students who attended the intervention (Attend) and those who did not (noAttend) were grouped further into at-risk and no-risk students. The Kruskal-Wallis test with Dunn's post-hoc was used to compare SAQ performance between the four groups. Data presented as median with IQR. N(at-risk): Attend = 78, noAttend = 115; N(no-risk): Attend = 211, noAttend = 213. *p<0.05 **p<0.01 *** p<0.001

Students perceive the activity to be helpful for their learning and preparation of the final exam

A total of 257 students responded to the following questions (Figure 1):

> What did you think of the activity that you have just completed?
> Do you consider it useful for your learning and exam preparation?
> Explain in a few lines.

Key themes and sub-themes from student responses are summarised in Figure 6. All following references to '%' are to the percentage of the 257 responses received. Themes and sub-themes are also not mutually exclusive. Briefly, most students thought the activity helped with their exam preparation (96%) by describing features of good SAQ answers (69%), and strategies on how to efficiently and effectively construct answers that adequately addresses the SAQ in the exam (30%).

Examples of student responses:

(1) "*I did think it was useful – gave an indication of depth needed in exam answers - gives an idea of how much study is required*"

(2)"*Yes it gives a good insight on what is expected and the different ways to answer questions*"

The activity managed to motivate 7% of students to work harder while the activity elicited negative emotions in the remaining 2%. Example of student responses with positive experience include:

(1)"*Understand the detail required for an adequate response. Wake up call*"

(2)"*Yes, allows me to understand what /how to answer these questions in the exam. Also, gives me a wake up call about how little I know*"

Examples of students with negative experiences include:

(1)"*It helped me understand what is expected. But it is making me stress because the answers are very complex"*

(2)"*It scared me as I think the time frame allowed for each question wouldn't allow us to write such detailed answers - such asthose [sic] required for full marks. Also diagrams arent [sic] quick to draw either*"

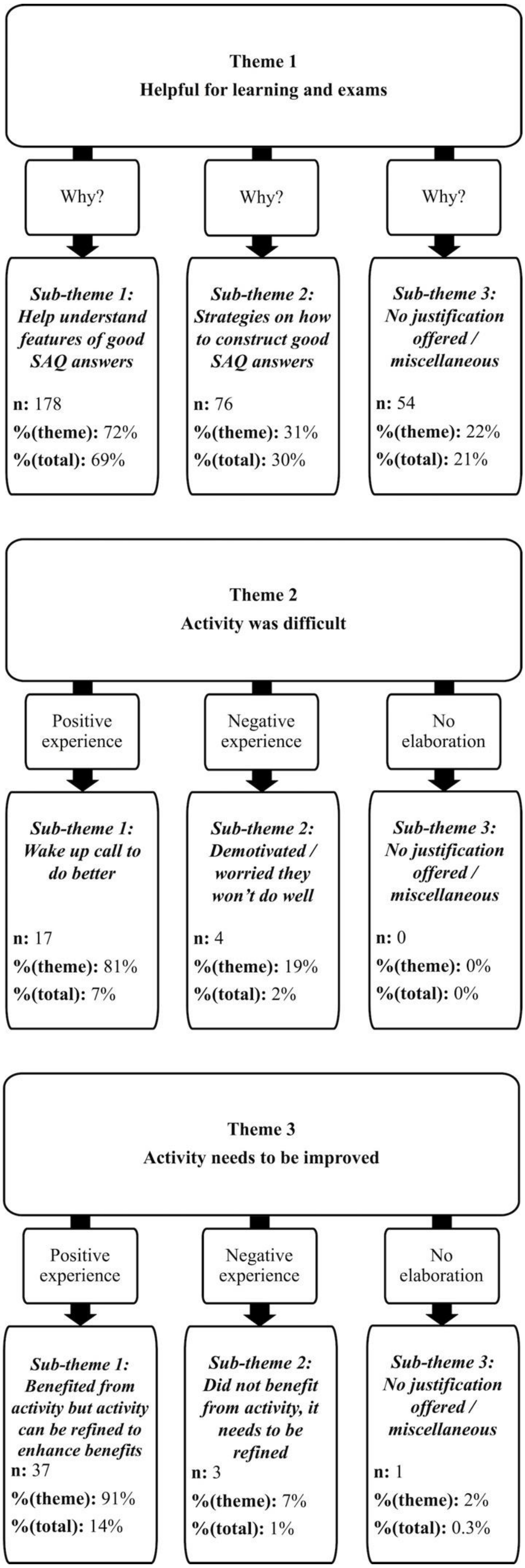

Theme 1
Helpful for learning and exams
Why?
Why?
Why?
Sub-theme 1: Help understand features of good SAQ answers
n: 178
%(theme): 72%
%(total): 69%
Sub-theme 2: Strategies on how to construct good SAQ answers
n: 76
%(theme): 31%
%(total): 30%
Sub-theme 3: No justification offered / miscellaneous
n: 54
%(theme): 22%
%(total): 21%
Theme 2
Activity was difficult
Positive experience
Negative experience
No elaboration
Sub-theme 1: Wake up call to do better
n: 17
%(theme): 81%
%(total): 7%
Sub-theme 2: Demotivated / worried they won't do well
n: 4
%(theme): 19%
%(total): 2%
Sub-theme 3: No justification offered / miscellaneous
n: 0
%(theme): 0%
%(total): 0%
Theme 3
Activity needs to be improved
Positive experience
Negative experience
No elaboration
Sub-theme 1: Benefited from activity but activity can be refined to enhance benefits
n: 37
%(theme): 91%
%(total): 14%
Sub-theme 2: Did not benefit from activity, it needs to be refined
n: 3
%(theme): 7%
%(total): 1%
Sub-theme 3: No justification offered / miscellaneous
n: 1
%(theme): 2%
%(total): 0.3%

Figure 6: Student perceptions of the intervention. Themes and sub-themes are not

mutually exclusive. %(theme) represents the proportion of students in the theme that fit into the respective sub-theme while %(total) indicates the proportion of the total number of students that is classified under the respective sub-theme. N(Total) = 257, N(Theme 1) = 248, N(Theme 2) = 21, N(Theme 3) = 41.

Discussion

LA-informed interventions are designed to remedy specific issues that impair student course performance and outcomes (Toetenel & Rienties, 2016; van Leeuwen et al., 2015). These targeted interventions are critical in higher education as factors informing student overall academic performance at universities are complex, and often filter down to the course level (Tinto, 2006). The intervention in our study was targeted towards improving student performance in the SAQ assessment task of a large, first year physiology course, by:

(1)improving student evaluative judgement (Carless & Chan, 2016; Carless et al., 2018; Tai et al., 2018; To & Carless, 2016)

(2)and facilitate adoption of learning strategies that beget enhanced task performance (e.g., conceptual integration across different physiological systems) (Biggs, 2003; Scouller, 1998).

This task was selected because it is a high stakes assessment task that is challenging for students in our course as evidenced by Appendix A and further highlighted in broader literature (Carnegie, 2015; Rozenblit & Keil, 2002; Sefton, 1998). The predictive models utilised in this study were essential in identifying low performing students who are at risk of failing the course and encourage their participation in the targeted intervention thereafter.

Student evaluative judgement of SAQ answer quality

Literature investigating approaches that are effective in improving student performance in the SAQ assessment task is limited, particularly in first-year university courses (Carnegie, 2015; Rashid-Doubell, O'Farrell, & Fredericks, 2018; Scoles, Huxham, & McArthur, 2013). Our findings and those of others suggest that providing students with example responses to a task, with instructor feedback, can improve their evaluative judgement and future task performance (Carless & Chan, 2016; Carless et al., 2018; Hendry et al., 2012; Hendry et al., 2011; Rust et al., 2003; Wimshurst & Manning, 2013). More importantly, unlike previous studies, we show that such improvements in understanding is possible in the first-year physiology course level, and can be facilitated in the absence of extensive discussions between students and instructors (e.g., Handley & Williams, 2011; Rashid-Doubell et al., 2018; Scoles et al., 2013; To & Carless, 2016). This could be because students in our study were exposed to a wider range of sample answers than that of previous studies and thus were exposed to a greater variety of standards and SAQ quality indicators, which would have had to be articulated via discussions between student and the instructor in previous studies (Sadler, 1989). Our findings also suggest that it is possible for other courses with a similar structure to ours (content-heavy, high student enrolments and contact hours) to articulate performance standards using examples in this manner without having to set aside precious time for extensive discussions between student and instructors.

Students' poor evaluative judgement prior to being given feedback could be attributed to their limited understanding of task requirements and associated notions of quality (Figures 3 and 4), as first-year university students are unfamiliar with the assessment practices of higher education (Kift & Moody, 2009; Yucel et al., 2014). Analysis of student self-scoring of their answer and examples provided insight to potential ill-fitted performance indicators responsible for the observed bias in student evaluative judgement prior to being given feedback. Doing so would allow us to better understand the standards used by naïve appraisers to evaluate work, which in turn permits the construction of feedback that is more targeted and effective in improving student evaluative judgement of similar work in the future.

(1) Effect of answer length on student evaluative judgement

Answer length was found to be one of the ill-fitted performance criteria that students used to evaluate the quality of SAQ answers, whereby students tend to assign a higher score for relatively lengthier examples, without considering the extent to which the content of the answer addresses task expectations. For instance, a large proportion of students underestimated short answers such as those with a diagram or two, while overestimating lengthier answers, such as those with substantial amounts of text (e.g., see answers scoring 2 vs. 1.5 in Q1 and answers scoring 3 vs. 2 in Q2, in Appendix B; Figure 4). However, in the students' view, a lengthier answer could be interpreted as one that included a greater number of concepts, which is indeed expected within a competent SAQ response (Winkielman, Schwarz, & Belli, 1998). Clearly, a lengthy SAQ answer is not necessarily a high-scoring one if the concepts included in the answer are not relevant to the question or if the length is a consequence of verbosity (e.g., answer with score 5 for Q2). Fortunately, qualitative findings from our study, especially comments such as "*Its very useful to help me what the answers culd [sic] reach the higher marks as short as possible. This time my answers are not very good for the correct answers*", suggest that feedback provided by the instructors, particularly those highlighting the importance of specificity and succinctness in good SAQ answers managed to reduce this bias in the student marking schema.

(2) Effect of question difficulty on student evaluative judgement

Students generally appear to experience greater difficulty in accurately self-evaluating their performance to more difficult questions that require complex answers, as evidenced by a higher variance in student self-score distributions for Q2 (Figure 3). Elevated answer complexity increases the number of ways in which answers of a particular standard can be expressed, which in turn limits students' capacity to scaffold their answer to a particular performance standard (Rozenblit & Keil, 2002). This uncertainty in performance assessment is compounded by the fact that key features of high-scoring answers, such as conceptual depth and integration, are abstract and difficult to apply when judging performance (Rust et al., 2003). However, decreased score variability in students' post-scores to Q2 suggests that using annotated examples to contextualise these criteria at different performance levels can mitigate this ambivalence.

Effect of intervention on student performance

Low and high performing students who attended the intervention performed significantly better in the SAQs of the EOS exam than their non-attending counterparts (Figure 5). It is clear that developing student understanding of task expectations and quality indicators as well as providing them with approximate guidelines on how to achieve them, can improve their performances in future tasks. This is because a clearer understanding of the features of a high-scoring answer to an SAQ provides students with more accurate learning goals, which in turn facilitates the adoption of learning strategies that are more appropriate to achieving those goals (Biggs, 2003; Scouller, 1998). Further, these students would also have a more refined evaluative judgement, which ensures that they are better equipped to identifying and addressing gaps in their performance as they work towards these goals (Tai et al., 2018). It is difficult to surmise that improvements in SAQ and course performance are solely due to the intervention. For instance, activity participants may comprise primarily of students with greater drive to improve their academic standing, as participation was not compulsory. Nevertheless, student perceptions of the activity were generally positive, which includes students' advocating its usefulness in supporting their learning (Figure 6) and enhanced evaluative judgement after feedback (Figure 4), even amongst at-risk students.

Low performing, at-risk students who participated in the intervention were still unable to match their SAQ as well as overall course performance with high, no-risk students who attended, and did not attend the intervention. This suggests that low performers require additional practice and discourse with instructors to further develop their understanding of critical but abstract criteria – e.g., demonstration of conceptual integration and depth (Rozenblit & Keil, 2002; Rust et al., 2003). Feedback in subsequent activity iterations can be developed to ensure that it is detailed, specific and directed towards describing how these abstract criteria is exemplified in text and how to demonstrate them in students' own work (Colthorpe, Liang, & Zimbardi, 2013). Alternatively, these students may require further interventional action to address other performance-limiting factors, uncaptured by our current study, as student poor performance is consequential of various factors (e.g., Code, 2020; Gašević et al., 2016). For instance, a lower SAQ and overall course performance suggests that at-risk students lack sufficient agency, or the capacity to engage and sustain productive academic behaviours to achieve high academic standards/goals (Code, 2020).

Conclusions and future directions

Findings from this study demonstrate how the LA paradigm can be used to inform effective intervention design, which targets specific issues that negatively affect student course performance. This is particularly apt for instructors in the current university climate where the student population is becoming increasingly more diverse and with it, the problems hindering student performance. Additionally, it is beneficial to target aspects that students will face in future years, such as answering SAQs in exams, as this impacts their overall academic performance at university.

Future studies can explore the applicability of this feedback paradigm in other assignments and disciplines as well as the extent to which the outcomes generated by this feedback modality differs to the conventional example and instructor-led discussion model. For instance, the activity can be incorporated into second- and third-year university courses our students would enrol into in the future. This would then permit students to progressively refine their evaluative judgement and SAQ performance by providing them with greater opportunities to practice and receive feedback. Furthermore, the intervention, which is now part of the course curriculum, can be refined in the future by supplementing it with a wider range of interventions, including those aimed at improving student agency, such as in Rathner, Hughes and Schuijers (2013).

References

Biggs, J. (2003). *Teaching for quality learning at university : What the student does* (2nd ed.). The Society for Research into Higher Education and Open University Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Carless, D., & Chan, K. K. H. (2016). Managing dialogic use of exemplars. *Assessment and Evaluation in Higher Education, 42*(6), 930-941. https://doi.org/10.1080/02602938.2016.1211246

Carless, D., Ho Chan, K. K., To, J., Lo, M., & Barrett, E. (2018). Developing evaluative judgement in higher education: Assessment for knowing and producing quality work (1st ed.). Routledge.

Carnegie, J. (2015). Use of feedback-oriented online exercises to help physiology students construct well-organized answers to short-answer questions. *CBE—Life Sciences Education, 14*(3), 1-12. https://doi.org/10.1187/cbe.14-08-0132

Code, J. (2020). Agency for learning: Intention, motivation, self-efficacy and self-regulation. *Frontiers in Education*, *5*(19). https://doi.org/10.3389/feduc.2020.00019

Colthorpe, K., Liang, S., & Zimbardi, K. (2013). Facilitating timely feedback in the biomedical sciences. *International Journal of Innovation in Science and Mathematics Education*, *21*(3), 60-74.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Sage.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*(C), 68-84. https://doi.org/10.1016/j.iheduc.2015.10.002

Handley, K., & Williams, L. (2011). From copying to learning: Using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education*, *36*(1), 95-108. https://doi.org/10.1080/02602930903201669

Harrell, J. F. E. (2015). *Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). Springer International Publishing.

Hazel, E., Prosser, M., & Trigwell, K. (2002). Variation in learning orchestration in university biology courses. *International Journal of Science Education*, *24*(7), 737-751. https://doi.org/10.1080/09500690110098886

Hendry, G. D., Armstrong, S., & Bromberger, N. (2012). Implementing standards-based assessment effectively: Incorporating discussion of exemplars into classroom teaching. *Assessment & Evaluation in Higher Education*, *37*(2), 149-161. https://doi.org/10.1080/02602938.2010.515014

Hendry, G. D., Bromberger, N., & Armstrong, S. (2011). Constructive guidance and feedback for learning: The usefulness of exemplars, marking sheets and different types of feedback in a first year law subject. *Assessment & Evaluation in Higher Education*, *36*(1), 1-11. https://doi.org/10.1080/02602930903128904

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Wiley. https://doi.org/ https://doi.org/10.1002/0471722146

Kift, S., & Moody, K. (2009). Harnessing assessment and feedback in the first year to support learning success, engagement and retention. In J. Milton, H. Cathy, J. Lang, G. Allan, & M. Nomikoudis (Eds.), *ATN Assessment conference 2009: Assessment in Different Dimensions* (pp. 1-12). Learning and Teaching Unit, RMIT University.

Kitsantas, A., & Zimmerman, B. (2006). Enhancing self-regulation of practice: The influence of graphing and self-evaluative standards. *Metacognition and Learning*, *1*(3), 201-212. https://doi.org/10.1007/s11409-006-9000-7

Megaw, P. L., & Zimanyi, M. A. (2019). Redesigning first year anatomy and physiology subjects for allied health students: Introducing active learning experiences for physiology in a first semester subject. *International Journal of Innovation in Science and Mathematics Education*, *27*(8), 26-35.

Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, *27*(4), 309-323. https://doi.org/10.1080/0260293022000001337

Rashid-Doubell, F., O'Farrell, P. A., & Fredericks, S. (2018). The use of exemplars and student discussion to improve performance in constructed-response assessments. *International journal of medical education*, *9*, 226-228. https://doi.org/10.5116/ijme.5b77.1bf6

Rathner, J., A., Hughes, D., L., & Schuijers, J., A. (2013). Redesigning a core first year physiology subject in allied health to achieve better learning outcomes. *International Journal of Innovation in Science and Mathematics Education*, *21*(2), 37-52.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*(5), 521-562. https://doi.org/10.1207/s15516709cog2605_1

Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*, *28*(2), 147-164. https://doi.org/10.1080/02602930301671

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119-144. https://doi.org/10.1007/BF00117714

Scoles, J., Huxham, M., & McArthur, J. (2013). No longer exempt from good practice: Using exemplars to close the feedback gap for exams. *Assessment & Evaluation in Higher Education*, *38*(6), 631-645. https://doi.org/10.1080/02602938.2012.674485

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *The International Journal of Higher Education and Educational Planning*, *35*(4), 453-472. https://doi.org/10.1023/A:1003196224280

Sefton, A. J. (1998). The future of teaching physiology: An international viewpoint. *Advances in Physiology Education*, *20*(1), S53-S58. https://doi.org/https://doi.org/10.1152/advances.1998.275.6.S53

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher education*, *76*(3), 467-481. https://doi.org/10.1007/s10734-017-0220-3

Tangalakis, K., Best, G., & Hryciw, D. H. (2017). Peer assisted study sessions for the development of transferable skills in undergraduate students from low socioeconomic backgrounds. *International Journal of Innovation in Science and Mathematics Education*, *25*(3), 36-44.

Tinto, V. (2006). Research and practice of student retention: What next? *Journal of College Student Retention: Research, Theory & Practice*, *8*(1), 1-19. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3583222/pdf/nihms438738.pdf

To, J., & Carless, D. (2016). Making productive use of exemplars: Peer discussion and teacher guidance for positive transfer of strategies. *Journal of Further and Higher Education*, *40*(6), 746-764. https://doi.org/10.1080/0309877X.2015.1014317

Toetenel, L., & Rienties, B. (2016). Learning design - creative design to visualise learning activities. *Open Learning: The Journal of Open, Distance and e-Learning*, *31*(3), 233-244. https://doi.org/10.1080/02680513.2016.1213626

van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Computers & Education*, *90*, 80-94. https://doi.org/https://doi.org/10.1016/j.compedu.2015.09.006

Vermunt, J. D. (2005). Relations between student learning patterns and personal and contextual factors and academic performance. *Higher education*, *49*(3), 205-234. https://doi.org/https://doi.org/10.1007/s10734-004-6664-2

Williams, M. T., Lluka, L. J., Meyer, J. H., & Chunduri, P. (2019). SOLO-based task to improve self-evaluation and capacity to integrate concepts in first-year physiology students. *Advances in Physiology Education*, *43*(4), 486-494. https://doi.org/10.1152/advan.00040.2019

Wimshurst, K., & Manning, M. (2013). Feed-forward assessment, exemplars and peer marking: Evidence of efficacy. *Assessment & Evaluation in Higher Education*, *38*(4), 451-465. https://doi.org/10.1080/02602938.2011.646236

Winkielman, P., Schwarz, N., & Belli, R. F. (1998). The role of ease of retrieval and attribution in memory judgments: Judging your memory as worse despite recalling more events. *Psychological Science*, *9*(2), 124-126. https://doi.org/10.1111/1467-9280.00022

Yucel, R., Bird, F. L., Young, J., & Blanksby, T. (2014). The road to self-assessment: Exemplar marking before peer review develops first-year students' capacity to judge the quality of a scientific report. *Assessment and Evaluation in Higher Education*, *39*(8), 971-986. https://doi.org/10.1080/02602938.2014.880400