SOLO-based task to improve self-evaluation and capacity to integrate concepts in first-year physiology students

Mark T. Williams,1 Lesley J. Lluka,1 Jan H. F. Meyer,2 and Prasad Chunduri1

1School of Biomedical Sciences, The University of Queensland, Brisbane, Queensland, Australia; and 2School of Civil Engineering, The University of Queensland, Brisbane, Queensland, Australia

Submitted 18 March 2019; accepted in final form 11 August 2019

Conceptual integration is critical in developing a deeper understanding of biological systems, such as the human body, as these systems are often coordinated and integrated entities (25). However, at tertiary institutions, physiology courses are traditionally taught on a module-by-module basis, where students learn about body systems, often in isolation from each other (23). Furthermore, students who are expected to integrate the various, separately taught, concepts are less inclined to do so when content is delivered in this manner, particularly in content-heavy introductory courses (11).

It is well known that variations in the complexity of student understanding exist (e.g., Ref. 13), which can be partly attributed to how students learn (1). Student learning, in turn, can be influenced by how well students are able to self-regulate the process (4). Through the years, various models have been constructed to define and subsequently measure the self-regulated learning process (4, 26, 30). However, generally, all models agree that this process involves students engaging in various strategies to attain specific goals or standards, while also monitoring their progress toward attaining them. It has been argued previously that the capacity to evaluate one’s performance accurately is one of the key determinants of a successful self-regulated learning process (26). However, studies have also clearly shown that inaccuracies in student capacity to evaluate their own performance exist (15, 17, 28). Generally, low-performing (Low) students tend to overestimate their own performance, whereas high-performing (High) students are inclined to underestimate themselves.

Increasing student self-evaluation accuracy by reducing the discrepancy between students’ perceived and actual ability can improve performance (14, 16, 20). Accuracy of student self-evaluation, in turn, can be enhanced through feedback from an external source, such as course instructors or peers (8). Empirical studies have indicated that feedback will need to highlight criteria and standards that are critical in demonstrating competence and how to achieve them in order for it to be effective in improving student self-evaluative abilities (6, 14, 21). However, it was argued that simply providing students with information on what is necessary to attain competence would allow students to develop a skill to evaluate the product, but not their ability to produce it (21).

Improvements in student ability in constructing the product, or complex understanding in this instance, could be achieved by guiding students through different stages of product construction and conveying key features of each stage (14, 21). However, it is not plausible to have multiple cycles of assessing student cumulative understanding of course content and providing personalized feedback at each instance. Instead, students in this present study were provided with standards, corresponding to the different levels of understanding, simultaneously. Findings from a recent study in a third-year-level civil engineering course support the validity of this design (19). Similar to Meyer et al. (19), the present study used the structure of the observed learning outcome (SOLO) taxonomy to con-
Struct responses that vary in structural complexity. This deviates from the conventional instructor-focused use of the SOLO taxonomy to assess student understanding of a topic (e.g., Refs. 7, 18). The SOLO taxonomy utilizes general, decontextualized descriptors that are grouped into five tiered categories (2), which adequately describe how understanding of topics in physiology develops. Briefly, individuals transition from the prestructural, unistructural, and multistructural stages as the number of concepts they know related to the topic increase (2). Integration of the knowledge from these concepts signifies the transition into the relational stage, and finally the extended abstraction stage is achieved when individuals are able to generalize conceptual knowledge across different contexts (2).

Although the present study applied the SOLO taxonomy in a manner similar to Meyer et al. (19), it also has some very important distinctions. First, to the best of our knowledge, the present study is the first to use SOLO in this manner in the field of physiology. Second, the majority of participants in the present study were first-year university students, who may be less adept at evaluating their own performance (10, 24) due to having less domain knowledge (15) and exposure to feedback (8, 20) than more advanced students. Furthermore, structural variations in this present study were framed using a topic with a relatively greater number of concepts and the extent to which these concepts are integrated with each other than in the previous study (see Ref. 19 and Supplemental Table S1 for a clearer comparison; all supplemental material is available at https://www.doi.org/10.5281/zenodo.3381721). For instance, answers in this present study were constructed from concepts pertaining to different course modules, such as cell signaling in the nervous and endocrine systems, as well as processes in the cardiovascular and respiratory systems (see Supplemental Table S1). In contrast, the answers from each workshop in Meyer et al. (19) focused on a specific concept, such as “critical flow.”

The present study aimed to evaluate the efficacy of exposing students to variations in structural complexity in facilitating their capacity to accurately self-assess their understanding. The present study also attempted to establish whether exposing students to these variations could translate to improved performance, specifically their ability to generate complex responses. Finally, the study examined whether the activity is able to replicate both of these outcomes in both the Low and High students.

METHODS

Course context. The present study was conducted in a large first-year physiology course at a large research-intensive university in Australia. Ethics approval for this project was obtained from The University of Queensland Human Behavioral and Social Sciences Ethical Review Committee. The course cohort consists mainly of first-year students, and the number of enrolled students varies from 700 to 1,000 every semester. This course covers and emphasizes the integration of various key biological processes in different physiological systems that are critical for complex organisms, particularly humans, to function. Students are provided with two lectures each week, along with an active-learning workshop to augment understanding of the concepts taught in lectures. Student attendance to lecture and workshop sessions, while not compulsory, was highly encouraged. The course is modular in nature and articulates knowledge of fundamental physiological processes, such as cell biology, nervous systems, support and movement, endocrinology, as well as circulation and gas exchange, to students. The course assessment items are used to gauge students’ understanding of the different biological concepts and their ability to integrate them, including progressive quizzes during the semester and the end-of-semester (EOS) exam.

The EOS exam consists of two components: multiple-choice questions and short-answer questions (SAQs). The SAQs are constructed in a manner that requires students to integrate concepts across different modules to perform well in the task. These SAQs generally start with a scenario description, followed by several subpart questions related to the scenario. For each question, marks are allocated to relevant key concepts and linkages, as determined by content experts responsible for constructing these questions. Examples of some subparts of integrative SAQs from the EOS exam in this particular course are provided below:

1. Which component of the nervous system controls the changes in the cardiovascular system during exercise?
2. Explain how the circulatory system and nervous system work together to increase the blood flow to the skeletal muscle by 25-fold during exercise.

The SOLO-based activity. The SOLO-based activity that forms the basis of the present study was conducted once per semester: during the active learning workshop session in the last week of the teaching semester. The present study reports on the data collected from the activities conducted in years 2014 and 2015.

At the start of the activity, students were required to answer an open-ended question on the concept of substance movement across the cell membrane. This was to expose students to their current understanding on this particular topic, as well as a means to self-evaluate. The question was:

Explain how molecules of different structures cross cell membranes and the functional importance of these processes in biology. You may include one or more diagrams in your answer. NOTE: Your answer will need you to draw on material across the various modules that you have covered in the course.

Students were then provided with nine model answers to the question (answers presented in the Supplemental Table S1), where each answer varied in structural complexity. Answers were randomized before the activity. Students were tasked to mark these answers based on their preconceptions of what features correspond to each level of structural complexity. They had to assign each model answer a unique score from 0 to 10, to represent the worst to the best answer. Each answer had to be sufficiently distinct and should capture key variations in structural complexity that corresponds to the different levels of student understanding. Therefore, an instructor, who is responsible for teaching this topic and for marking the related content in the exam and has over 30 yr of tertiary teaching experience, was tasked with constructing these answers. The co-coordinator of the course, also a course content instructor whose expertise lies in a related field of physiology (the endocrine and cardiovascular system), with over 10 yr of experience, then reviewed the answers. They then engaged in discourse and agreed on the structure and content of the answers before implementing them in the activity.

The subsequent section essentially represents the intervention phase of the activity, which is where students were provided with the instructor-assigned score for each model answer with an explanation to justify the scoring scheme. Following this, students were tasked to evaluate their own answer and provide a mark, once again out of 10. Students were also asked to justify their marking by selecting one of the five statements, derived from the SOLO taxonomy, that best described the order of complexity of their answer (Table 1). The instructor then evaluated each student response in the same manner after the activity has concluded. To enhance effectiveness of activity, students were permitted to take the model answers with instructor scores home, for further review in their own time.

Since 2015, an additional section was incorporated into the activity structure to allow us to assess the effects of the model answers on student self-evaluation accuracy. Instead of just self-evaluating them-
Defining the various student groups in the present study. Student data were de-identified before any analysis. An overview of the various student groups and how these groups contribute to addressing the present study’s aims is provided in Fig. 1.

Evaluating student capacity to distinguish between different levels of complexity. The capacity to distinguish between the different levels of understanding is one of the key factors necessary to permit accurate self-evaluation of student understanding. Thus student ability to differentiate between the model answers of different complexity levels, before the exposure to instructor marking scheme, was studied. This was achieved by evaluating whether student scoring of each of the model answers, except the best answer, was significantly different from that of the ideal instructor score for both cohorts. For the best answer, the proportion of students who matched the instructor score was determined and compared between the best and the remaining model answers. Student marking of the best answer was analyzed differently because it is not possible to assign a score greater than 10 to the best answer (as per the instructions given to students during the activity), which violates the statistical test assumption, as detailed in the statistical analysis section below.

Evaluating the effect of SOLO-based activity on student capacity to self-evaluate. The difference between student and instructor marking of student responses to the open-ended question was used to gauge student ability to self-evaluate their understanding. A greater difference implies a lower capacity to self-evaluate. Justification statements were ordered from 1 to 5, in the order of prestructural to extended abstract. The absolute difference between student- and instructor-marking scores of the best and remaining model answers was compared.

Table 1. Statements and the corresponding SOLO levels that were utilized in marking student responses to the open-ended questions of the activity

| Statement No. | Statement Adapted Structurally from the SOLO Taxonomy | SOLO Level |
|---------------|------------------------------------------------------|------------|
| I             | There is no indication of known relevant knowledge or relevant connection between the question and the answer; the question is reflected back in the answer. | Prestructural |
| II            | The answer only considers one relevant aspect connected to the question | Unistructural |
| III           | The answer contains several aspects, but these are conceptually independent. | Multistructural |
| IV            | The answer contains several conceptually related aspects that provide a generic answer that covers all context | Relational |
| V             | The answer contains related and relevant aspects plus hypotheses, generalizing beyond the information contained in the question to other contexts. | Extended abstraction |

See Biggs and Collis (2). SOLO, structure of the observed learning outcome.

Fig. 1. Flow chart depicting the student groups employed to address each aim of the present study. Study aims are depicted in boxes with a red outline, and red arrows indicate which student groups were used to address the respective aims. There were 944 students who attended the workshop for structure of the observed learning outcome (wSOLO) and 743 who did not attend the SOLO workshop (noSOLO) in the 2014 cohort. The 2015 cohort had 824 wSOLO and 806 noSOLO students. The no. of responses that fulfilled the selection criteria were 309 (2014) and 355 (2015). The median score was 4 for the 2014 cohort and 3 for the 2015 cohort. There were 191 low-performing students and 118 high-performing students in the 2014 cohort, whereas the 2015 cohort had 215 low-performing students and 140 high-performing students. Greater details of each evaluation are presented in the text.
assigned score and selected justification statement was calculated for each student in the 2015 cohort, before and after the intervention section of the activity. As 2014 students were not asked to assess their own responses before the intervention, the difference statistic was only calculated postintervention. All subsequent analyses that compared changes in student self-evaluation as a consequence of exposing them to the model answers and instructor marking schema were conducted using the 2015 cohort.

The difference statistic was specifically used to determine whether:

1. Variations and instructor feedback altered self-evaluation of all of the participating students, within both Low and High students, by comparing difference statistics postintervention to preintervention.
2. There is a difference in alteration of self-evaluation between Low and High students, by comparing changes in self-evaluation of the two groups.
3. Changes in self-evaluation are maintained across multiple iterations of the activity by comparing the difference statistic postintervention of the 2015 cohort to the 2014 cohort.

**Evaluating the efficacy of SOLO-based activity in improving student ability to integrate concepts.** Student performance in the SAQs of the EOS exam was used as a measure of student ability to integrate biological concepts to generate complex responses. For both the 2014 and 2015 cohorts, the total score obtained by each student in the SAQs was calculated. SAQ scores of students who participated in the SOLO-based activity (wSOLO) were then compared with those who did not (noSOLO). Differences in SAQs scores between Low, High, and noSOLO students within each cohort were also tested for significance. First attempt scores were used.

**Statistical analysis.** Statistical analysis was conducted using the GraphPad Prism software (version 6), and the threshold to achieve statistical significance was set to \( P < 0.05 \). The Wilcoxon signed-rank test was used to determine whether student scoring of each model answer, except the best model answer, was significantly distinct from the ideal instructor score. The signed-rank test was not used to evaluate the accuracy of student scoring of the best model answers, as instructor- and student-assigned scores can never be distributed symmetrically around the median score of 10 (best answer). Instead, Fisher’s exact test was used to evaluate whether there is a difference in student proportions that correctly scored the best model answer and the remaining model answers as a pseudo-measure of marking accuracy of the best answer.

The Wilcoxon matched-pairs signed-rank test was used to determine whether the absolute difference statistic, for both scores and SOLO statements, was significantly different after exposure to variations in structural complexity. The Mann–Whitney test was used to compare difference statistics for both scores and SOLO statements:

1. Of student self-marking, postintervention of the 2014 cohort to the 2015 cohort.
2. Between Low and High students of the 2015 cohort, before and after being exposed to variations in structural complexity.

The difference in wSOLO and noSOLO student performance in the SAQs of the EOS exam was compared using the Mann–Whitney test. Low, High, and noSOLO student performance in the SAQs was compared using a Kruskal–Wallis test with Dunn’s post hoc analysis.

**RESULTS**

**Student ability to distinguish between different complexity levels.** Student-assigned scores for both 2014 and 2015 cohorts were found to spread over a large range for model answers 2–9 in both cohorts (Fig. 2), although the median student-assigned scores for both cohorts was shown to generally increase, along with the complexity of the model answers. The median student scoring of every model answer, except answer 6 for the 2014 cohort, was significantly different from the ideal instructor score (answer 2: \( P = 0.02 \); answer 7: \( P = 0.002 \), remaining significant differences: \( P < 0.001 \); Fig. 2A). The 2015 student scoring for every model answer, except answer 7, was found to be significantly different from the instructor score (\( P < 0.001 \) for every significant difference; Fig. 2B).

Less than 50% of students in the 2014 and 2015 cohort were able to correctly score every model answer (Fig. 3). Variations in marking accuracy exist between the 2014 (Fig. 3A) and 2015 (Fig. 3B) cohorts. Model answer 10 (best answer) had the greatest proportion of students with the correct score assigned to it (2014, answer 2: \( P = 0.04 \), answers 3–7: \( P < 0.001 \), answer 8: \( P = 0.42 \), answer 9: \( P = 0.11 \); 2015, answer 2: \( P = \)

![Fig. 2. Distribution of assigned scores for each model answer for students from the 2014 (A) and the 2015 cohort (B). Horizontal lines represent the median student-assigned score for each model answer and the interquartile range. Model answers are ordered based on the score assigned to the model answers by the instructor, with higher scores for more complex answers. Model answer 10 was not included in the analysis, as it is not possible for student and instructor scores to be distributed above 10. The Wilcoxon signed-rank test was used to determine whether student-assigned scores were different from the instructor score for each model answer. Total no. of students (n) for each model answer was 309 for the 2014 cohort and 355 for the 2015 cohort. \( P < 0.05 \) was used as a threshold for significance. *** \( P < 0.001 \), ** \( P < 0.01 \), * \( P < 0.05 \).](http://advan.physiology.org)
Changes in student self-evaluation. Student marking of their own answer aligned closer to the instructor marking postintervention (Fig. 4). This applies for both assigned scores ($P < 0.001$, Fig. 4A) and SOLO statements ($P < 0.001$, Fig. 4B), although this decline was not as pronounced for SOLO statements. The extent to which student self-assigned scores deviated from the instructor postintervention was not significant between the 2014 and 2015 cohort ($P = 0.47$, Fig. 4C).

Fig. 3. Proportion of student-assigned scores that matched (open bar), was greater than (shaded bar), or was lower than (solid bar) that of the instructor, for each model answer. A: results from the 2014 cohort. B: results from the 2015 cohort. Model answers are ordered based on the score assigned to the model answers by the instructor, with higher scores for more complex answers. The percentage of the total no. of students who were present in each of these cohorts for each model answer is overlaid on the respective bars. Fisher’s exact test was used to determine if the proportion of matched students in answer 10 was significantly different from that of the remaining model answers. Total no. of students ($n$) for each column was 309 for 2014 cohort and 355 for 2015 cohort. $P < 0.05$ was used as a threshold for significance. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.

0.009, answer 3: $P = 0.54$, answers 4–8: $P < 0.001$, answer 9: $P = 0.21$; Fig. 3). Note, although model answer 3 had the highest proportion of 2015 students with the correct score, it was not significantly higher than that of answer 10.

Changes in student self-evaluation. Student marking of their own answer aligned closer to the instructor marking postintervention (Fig. 4). This applies for both assigned scores ($P < 0.001$, Fig. 4A) and SOLO statements ($P < 0.001$, Fig. 4B), although this decline was not as pronounced for SOLO statements. The extent to which student self-assigned scores deviated from the instructor postintervention was not significant between the 2014 and 2015 cohort ($P = 0.47$, Fig. 4C).

Fig. 4. Alterations in student marking of his/her own answer to the open-ended question following exposure to model answers that varied in structural complexity. Bars represent the median absolute difference between student- and instructor-assigned scores (A and C) and structure of the observed learning outcome (SOLO) statements (B and D) for student answers to the open-ended question. The error bars depict the interquartile range. A and B: deviations between marking schemes before (Pre) as well as after exposure (Post) to the model answers of students in the 2015 cohort. C and D: the differences in marking schemes were compared between the 2014 and 2015 cohort. Wilcoxon matched-pair signed-rank test was used to compare between groups in A and B, whereas the Mann–Whitney test was used to compare cohorts in C and D. No clear differences in distribution between the two groups can be observed in B, but test statistics, Wilcoxon (Pre vs. Post) SOLO = −6.667, suggest that points in Post group are distributed closer to 0. 2014 cohort: $n = 309$ students; 2015 cohort: $n = 355$ students. $P < 0.05$ was used as a threshold for significance. ***$P < 0.001$. 

490 SOLO-BASED TASK FOR FIRST-YEAR PHYSIOLOGY STUDENTS

Advances in Physiology Education • doi:10.1152/advan.00040.2019 • http://advan.physiology.org
However, misalignment between student- and instructor-assigned SOLO statements was greater in the 2015 cohort than the 2014 cohort ($P < 0.001$, Fig. 4D).

Scores and SOLO statements assigned by the Low student group to their own responses to the open-ended question were aligned closer to the instructor postintervention ($P < 0.001$ for all comparisons, Fig. 5). Conversely, a significant reduction in difference statistics was only observed for SOLO statements in High students (score: $P = 0.08$, Fig. 5A; SOLO: $P = 0.003$, Fig. 5B). Improvements in student marking schemes for both scores and SOLO statements were more pronounced in Low students. However, High students still had a significantly lower difference statistic, for both scores and SOLO statements, than Low students preintervention and postintervention ($P < 0.001$ for all comparisons, Fig. 5).

**Student ability to integrate concepts.** Students who attended the workshop, or wSOLO students, were found to have a significantly greater score in the SAQs than those who did not, or noSOLO students, across all four semesters ($P < 0.001$ for all comparison, Fig. 6, A and B). High students performed the best in the SAQs of the EOS exam followed by Low and finally noSOLO students in every semester ($P < 0.001$ for all comparisons, Fig. 6).

**DISCUSSION**

An accurate evaluation of students’ own performance is necessary for a successful self-regulated learning process (26), which results in effective learning (4) and enhanced understanding (1) and subsequently elevated performance in assessment tasks (16). However, students have been shown to be inaccurate in the self-evaluation of their own performance in the field of physiology, with greater inaccuracies observed in low performers (17, 28). Empirical evidence obtained from studies in different contexts suggests that supplementing students with standards corresponding to different levels of competence can improve student self-evaluation accuracy and performance (14, 19). The present study aimed to appraise the capacity of exposing students to variations in structural complexity in facilitating their aptitude to accurately assess the complexity of their own understanding. The present study also endeavored to establish if exposing students to these variations could ameliorate student ability to generate complex responses of higher marks to integrative SAQs in the EOS exam. Finally, the study examined whether the activity is able to replicate both of these outcomes in both the Low and High cohorts of students.

Exposing students to variations in structural complexity and associated instructor oral feedback improved student capacity to ascertain the complexity of their own understanding (score; Fig. 4A) and the criteria used to evaluate their competency (SOLO statements; Fig. 4B). This suggests that students were basing their self-evaluation on a set of internalized performance standards (4, 26, 30) that is relatively distinct from that of the instructor, before the intervention. The conjecture is further evidenced by the fact that student marking of most model answers was different from that of the instructor (Figs. 1 and 2). Thus it is evident here that feedback provided to students during this activity improved student self-evaluation accuracy by providing students with a set of “new standards,” representative of the various levels of conceptual integration, and thus understanding. It prompted students to revisit their existing internal standards and subsequently reconstruct standards that are better aligned with that of the instructor (3).

Refining student self-evaluation accuracy by supplementing students with criteria, which articulate features necessary to attain competence in a task, has been well established in

![Fig. 5](http://advan.physiology.org)

Fig. 5. Changes in low-performing (Low) and high-performing (High) student marking of his/her own answer to the open-ended question following exposure to model answers that varied in structural complexity. Bars represent the median absolute difference between both student groups and instructor-assigned scores (A) and structure of the observed learning outcome (SOLO) statements (B). The error bars depict the interquartile range. Only 2015 students were included in the analysis, as 2014 students did not evaluate their own answers before the intervention section of the activity. Students were grouped based on the score assigned by the instructor to their responses to the open-ended question of the activity. Low students had scores below the median instructor-assigned score, whereas High students scored equal to or above (median: 3). Deviations between marking schemes before (Pre) as well as after (Post) exposure to the model answers of students are presented as a shaded bar and open bar, respectively. Wilcoxon matched-pair signed-rank test was used to compare Pre to Post performance within Low and High groups in A and B, whereas the Mann–Whitney test was used to compare within Pre and Post performance between groups in A and B. $P < 0.05$ was used as a threshold for significance. ***$P < 0.001$, **$P < 0.01$. 

*Advances in Physiology Education* • doi:10.1152/advan.00040.2019 • [http://advan.physiology.org](http://advan.physiology.org)
literature. It has been shown that student self-evaluation accuracy of exam performance did not improve (5), unless supplemented with guidelines that they can utilize to evaluate their performance (6). Furthermore, it has been shown that development of self-evaluation accuracy can be amplified by facilitating the identification of a greater number of criteria associated with each performance standard. For instance, Bol et al. (6) discovered that students who received guidelines on self-evaluation, combined with group discussions and peer feedback about their self-evaluation, had greater self-evaluation accuracy than those who only received guidelines. Additionally, Orsmond et al. (21) observed improvements in self-evaluation of student work, specifically scientific posters, when marking criteria was supplemented with exemplars.

Students appear to be better at identifying the most complex model answer, relative to other answers, before any feedback from the content expert. This suggests that exposing students to variations in complexity alone has some merit in driving the discovery of features pertaining to complex understanding among them. Alternatively, this could mean that a vast number of students are already cognizant of features pertaining to complex understanding before the intervention and students having a relatively poor self-evaluation at the start of the activity is due to their inability to effectively discriminate between different complexity levels. However, it is difficult to supply strong evidence for either conjecture due to the manner in which the activity is structured. Nevertheless, the capacity to identify features necessary to demonstrate the most complex understanding of a concept alone is not sufficient to maximize improvements in student self-evaluation. Instead, an ability to distinguish between the different degrees of complexity would be beneficial, as discussed previously.

Segregating students based on their competency revealed that modifications in self-evaluation accuracy appear to be specific to student performance levels and the evaluation metric (i.e., scores and SOLO statements; Fig. 5). The intervention improved Low students’ capacity to identify the complexity of their understanding (scores) and the reason for their poor performance (statements). This is unsurprising, as Low students were shown to have poorer self-awareness than High students (5, 15, 17, 28), which is partly attributed to the lack of sufficient domain knowledge required to accurately identify gaps in their abilities (15, 29). This is reinforced by the fact that the intervention was not able to refine High students’ capacity to discern between different complexity levels (score; Fig. 5A). However, high performers appear to restructure their internal
In one study, participants demonstrated greater awareness of each stage. This is also exemplified in previous studies in ent stages of product construction, conveying key features of product could be achieved by guiding students through different processes, as conceptual integration is one of the key factors regulating their learning process by ensuring that learning is facilitated improvements in standards students utilize in the SAQs. This suggests that variations in structural complexity facilitated improvements in standards students utilize in regulating their learning process by ensuring that learning approaches employed are better tailored to achieve conceptual integration. Indeed, improved SAQ performance suggests that students were attempting to identify associations between concepts/modules taught in the course during their learning process. This indicates that the present activity is also capable of improving self-assessment in both novices, specifically first-year university students, as seen in our study, and in experienced self-evaluators, namely third-year university students in Meyer et al. (19). Neuroimaging and behavioral studies have suggested that individuals develop greater self-awareness of their cognitive processes and abilities as they age, with education further enhancing improvements in self-awareness. Indeed, education science research has indicated that students in their third year of university would be relatively more proficient in their capacity to self-evaluate. This could be due to third-year students having:

1. Greater domain knowledge, and, therefore, they are more aware of the standards necessary to achieve competence in the domain.
2. Received a greater amount of feedback on their learning, which would allow them to develop a greater awareness of their learning process.

Thus, as self-evaluation abilities improve while students’ progress through university, it is imperative then to ensure that activities such as this are implemented early in higher education to maximize the benefits they offer.

Students who attended the SOLO-based activity, regardless of their competency, constructed more complex responses to the SAQs. This suggests that variations in structural complexity facilitated improvements in standards students utilize in regulating their learning process by ensuring that learning approaches employed are better tailored to achieve conceptual integration. Indeed, improved SAQ performance suggests that students were attempting to identify associations between concepts/modules taught in the course during their learning process, as conceptual integration is one of the key factors required to perform well in this section of the exam. This validates the proposition offered by Orsmond et al. (21), who argues that improvements in student ability to construct a product could be achieved by guiding students through different stages of product construction, conveying key features of each stage. This is also exemplified in previous studies in which feedback was presented in a different manner. In one study, participants demonstrated greater awareness of their aptitude and changes associated with it when supplemented with finer performance goals than those who received the highest attainable standard as a goal after each session of testing. In another study, students receiving feedback focusing on the process used to construct an answer to a quiz question performed better in subsequent quizzes than students who only received feedback on the answer. Lastly, the fact that these SAQs were not specifically based on the concepts covered in the intervention also suggests that students were focused on identifying deficiencies in the structural complexity of their understanding more broadly and were not restricted to a certain concept.

Limitations and future directions. The present study provides preliminary evidence on the effectiveness of the SOLO-based activity in improving student ability to self-evaluate their understanding. However, it is difficult to ascertain the effectiveness of the feedback, framed in the manner seen in our study, in improving student self-evaluation relative to other feedback types. Such comparisons would, of course, serve as a compelling hypothesis to be tested in future iterations of the activity.

Low students were still underperforming in both their capacity to self-evaluate and in constructing complex responses compared with their High colleagues, at the conclusion of this study. A possible avenue for improvement is through providing additional feedback that is fine-tuned to each student, specifically by highlighting strengths, weaknesses, and the relative complexity of their own responses. This could be achieved by having guided discussions, as utilized in Bol et al. (6), of student responses and model answers and/or by returning marked student responses to the open-ended question postactivity. Studying student rationale underlying self-scoring before and after the intervention via open-ended questions would also facilitate the discovery of any weaknesses in standards guiding the learning process. Identified common variations and gaps in understanding, uncaptured by the present range of model answers, could also be incorporated into answers in future iterations of the activity.

In conclusion, our study has found some evidence that exposing students to variations in structural complexity underlying different levels of conceptual understanding can improve students’ ability to self-evaluate the complexity of their own understanding. This intervention also led to students generating improved responses to integrative SAQs in the exam, a measure of their capacity to integrate concepts across modules. These positive results, along with those in Meyer et al. (19), also affirm that the SOLO-based activity outcomes can be replicated in other disciplines in which knowledge is developed in the same manner.

DISCLOSURES

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

AUTHOR CONTRIBUTIONS

L.J.L., J.H.F.M., and P.C. conceived and designed research. M.T.W., L.J.L., J.H.F.M., and P.C. performed experiments; M.T.W. analyzed data; M.T.W. interpreted results of experiments; M.T.W. prepared figures; M.T.W. drafted manuscript; M.T.W., L.J.L., J.H.F.M., and P.C. edited and revised manuscript; M.T.W., L.J.L., J.H.F.M., and P.C. approved final version of manuscript.

Advances in Physiology Education • doi:10.1152/advan.00040.2019 • http://advan.physiology.org
REFERENCES

1. Biggs J. Individual differences in study processes and the quality of learning outcomes. *High Educ* 8: 381–394, 1979. doi:10.1007/BF01680526.

2. Biggs JB, Collis KF. *Evaluating the Quality of Learning: the SOLO Taxonomy (Structure of the Observed Learning Outcome).* New York: Academic, 1982.

3. Boekaerts M. Subjective competence, appraisals and self-assessment. *Learn Instr* 1: 1–17, 1991. doi:10.1016/0959-4752(91)90016-2.

4. Boekaerts M. Self-regulated learning at the junction of cognition and motivation. *Eur Psychol* 1: 100–112, 1996. doi:10.1016/1076-9998(96)90120-1.

5. Bol L, Hacker DJ, O'Shea P, Allen D. The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *J Exp Educ* 73: 269–290, 2005. doi:10.1080/00220670209596583.

6. Bol L, Hacker DJ, Walck CC, Nunnery JA. The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemp Educ Psychol* 37: 280–287, 2012. doi:10.1016/j.cedpsych.2012.02.004.

7. Brabrand C, Dahl B. Using the solo taxonomy to analyze competence progression of university science curricula. *High Educ* 58: 531–549, 2009. doi:10.1007/s10734-009-9210-4.

8. Butler DL, Winne PH. Feedback and self-regulated learning: a theoretical synthesis. *Rev Educ Res* 65: 245–281, 1995. doi:10.3102/00346543065003245.

9. Chan CC, Tsui MS, Chan MYC, Hong JH. Applying the structure of the observed learning outcomes (SOLO) taxonomy on student’s learning outcomes: an empirical study. *Asses Eval High Educ* 27: 511–527, 2002. doi:10.1080/026029302220002292.

10. Demetriou A, Bakracevic K. Reasoning and self-awareness from adolescence to middle age: organization and development as a function of education. *Learn Individ Differ* 19: 181–194, 2009. doi:10.1016/j.lindif.2008.10.007.

11. Ding L, Mollohan KN. How college-level introductory instruction can impact student epistemological beliefs. *J Coll Sci Teach* 44: 19–27, 2015. doi:10.25054/jcst15.044_04_19.

12. Falchikov N, Boud D. Student self-assessment in higher education: a meta-analysis. *Rev Educ Res* 65: 395–430, 1995. doi:10.3102/00346543065003245.

13. Hegarty-Hazel E, Prosser M. Relationship between students’ conceptual knowledge and study strategies. Part 2. Student learning in biology. *Int J Sci Educ* 13: 421–429, 1991. doi:10.1080/0950069910130405.

14. Kitsantas A, Zimmerman B. Enhancing self-regulation of practice: the influence of graphing and self-evaluative standards. *Metacogn Learn* 1: 201–212, 2007. doi:10.1007/s11409-006-9000-7.

15. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one’s own incompetence lead to inflated self-assessments. *J Pers Soc Psychol* 77: 1121–1134, 1999. doi:10.1037/0022-3514.77.6.1121.

16. Labuhn A, Zimmerman B, Hasselhorn M. Enhancing students’ self-regulation and mathematics performance: the influence of feedback and self-evaluative standards. *Metacogn Learn* 5: 173–194, 2010. doi:10.1007/s11409-010-9056-2.

17. Langendyk V. Not knowing that they do not know: self-assessment accuracy of third-year medical students. *Med Educ* 40: 173–179, 2006. doi:10.1111/j.1365-2929.2005.02372.x.

18. Leung CF. Assessment for learning: using solo taxonomy to measure design performance of design and technology students. *Int J Technol Des Educ* 10: 149–161, 2000. doi:10.1023/A:1008937007674.

19. Meyer JHF, Knight DB, Callaghan DP, Baldock TE. An empirical exploration of metacognitive assessment activities in a third-year civil engineering hydraulics course. *Eur J Eng Educ* 40: 309–327, 2015. doi:10.1080/03403797.2014.960367.

20. Niefeld JL, Schraw G. The effect of knowledge and strategy training on monitoring accuracy. *J Educ Res* 95: 131–142, 2002. doi:10.1080/00220670209596583.

21. Orsmond P, Merry S, Reiling K. The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assess Eval High Educ* 27: 309–323, 2002. doi:10.1080/026029302220001337.

22. Paulson Gjerde K, Padgett MY, Skinner D. The impact of process vs. outcome feedback on student performance and perceptions. *J Learn High Educ* 13: 73–82, 2017.

23. Rodríguez-Barbero A, López-Novoa JM. Teaching integrative physiology using the quantitative circulatory physiology model and case discussion method: evaluation of the learning experience. *Adv Physiol Educ* 32: 304–311, 2008. doi:10.1152/advan.00107.2008.

24. Sebastian C, Burnett S, Blakemore S-J. Development of the self-concept during adolescence. *Trends Cogn Sci* 12: 441–446, 2008. doi:10.1016/j.tics.2008.07.008.

25. Sefton AJ. The future of teaching physiology: an international viewpoint. *Am J Physiol* 275: S3–S58, 1998. doi:10.1152/advances.1998.275.6.S33.

26. Winne PH. Inherent details in self-regulated learning. *Educ Psychol* 30: 173–187, 1995. doi:10.1207/s15326985ep3004_2.

27. Winne PH, Hadwin AF. Studying as self-regulated learning. In: *Metacognition in Educational Theory and Practice*, edited by Hacker DJ, Dunlosky J, Graesser AC. Mahwah, NJ: Erlbaum, 1998, p. 27–30.

28. Ziegler B, Montplaisir L. Student perceived and determined knowledge of biology concepts in an upper-level biology course. *Adv Physiol Educ* 32: 304–311, 2008. doi:10.1152/advan.00107.2008.

29. Zimmerman BJ. Self-regulated learning and academic achievement: an overview. *Educ Psychol* 25: 3–17, 1990. doi:10.1207/s15326985sep2501_2.

30. Zimmerman BJ. Attaining self-regulation: a social cognitive perspective. In: *Handbook of Self-Regulation*, edited by Boekaerts M, Pintrich PR, Zeidner M. San Diego, CA: Academic, 2000, chapt 2, p. 13–39.