TEACHING INNOVATIONS

Incorporating higher order thinking and deep learning in a large, lecture-based human physiology course: can we do it?

Justine O. Hobbins, Coral L. Murrant, Laelie A. Snook, Justine M. Tishinsky, and Kerry L. Ritchie

Department of Human Health and Nutritional Science, University of Guelph, Guelph, Ontario, Canada

Submitted 10 September 2019; accepted in final form 20 August 2020

Hobbins JO, Murrant CL, Snook LA, Tishinsky JM, Ritchie KL. Incorporating higher order thinking and deep learning in a large, lecture-based human physiology course: can we do it? Adv Physiol Educ 44: 670–678, 2020; doi:10.1152/advan.00126.2019.—Large classes taught with didactic lectures and assessed with multiple-choice tests are commonly reported to promote lower order (LO) thinking and a surface approach (SA) to learning. Using a case study design, we hypothesized that incorporating instructional scaffolding of core physiology principles and assessing students exclusively with long-answer written tests would encourage higher order (HO) thinking and promote a deep approach (DA) to learning in a two-course physiology sequence (Phys I and II), despite their large size. Test questions were categorized as LO or HO according to the Blooming Biology Tool, and students’ LO and HO performance was determined for each of six tests across the two courses. The validated Revised Two-Factor Study Process Questionnaire survey tool was administered at the beginning and end of each course to measure student approach to learning. HO performance was maintained across Phys I (72.9 ± 19.4% vs. 74.8 ± 20.7%, P = 0.37) and significantly improved across Phys II (69.9 ± 18.4% vs. 79.4 ± 14.8%, P < 0.001). Unexpectedly, students’ LO performance declined from the beginning to end of Phys I (78.5 ± 20.6% vs. 69.4 ± 17.9%, P < 0.001) and Phys II (80.5 ± 19.6% vs. 72.2 ± 24.3%, P < 0.001). Students’ approach to learning did not change throughout Phys I or II, but at each time point students preferred a DA over a SA. Taken together, these results indicate that an intentionally designed large lecture class can support a DA to learning and suggests that this teaching and assessment structure may be particularly well suited to promote HO thinking, albeit possibly at the expense of LO thinking.

INTRODUCTION

Disciplinary content, learning objectives, and available resources must be considered when designing teaching and assessment methods for a given classroom. Physiology is based on the integration of, and relationships between, systems to solve problems and, therefore, requires students to work largely at higher levels of Bloom’s Taxonomy (10). Indeed, disciplinary content knowledge and understanding of core principles (e.g., homeostasis, information flow) (24) in physiology is important [i.e., lower order (LO) thinking]. However, learners must also recognize that these core principles are not completely distinct from one another and be able to see relationships between systems, using this information to solve problems and predict outcomes [i.e., higher order (HO) thinking] (24, 27).

Due to the constraints placed on teaching and assessment methods by class size (35), the traditional lecture with an instructor focus is the most common approach adopted by professors in large university classes (11, 31, 42). Often associated with traditional lectures are didactic teaching methods and static PowerPoint presentations (20), where students are more passive rather than active listeners (9, 26). Under this method of instruction, students often adopt a surface level approach to learning, memorizing disconnected facts and relying on rote problem-solving methods (42, 49), utilizing LO levels of understanding to solve problems. As modifications to teaching methods are made, variations in the implementation of traditional lecture methods result, creating a spectrum of instructor-centered to student-centered teaching methods. For example, at the opposite end of the spectrum to traditional lectures may be case-based, problem-based, and peer learning, which have been shown to foster deep learning (7, 41) and to lead students to HO levels of understanding involving critical thinking (7, 9) (Fig. 1). While students utilizing a surface approach (SA) to learning are often operating at a lower cognitive level, when students adopt a deep approach (DA) to learning, they effectively utilize a range of thinking skills spanning all levels of Bloom’s Taxonomy (7); they can recall information about a subject (level 1: knowledge, LO), but are also able to move up through the hierarchical levels to describe the concept (level 2: comprehension, LO), apply it to new situations (level 3: application, HO), understand how it relates to other concepts as a whole (level 4: analysis, HO), and use it to create (level 5: synthesis, HO) and critique (level 6: evaluation, HO) (7). However, despite the apparent advantage of more active, student-centered approach to teaching, resource demands (time, personnel, financial) associated with restructuring a large course to include these methods are commonly noted as barriers for implementation by instructors (42, 44). It is also worth noting that, while these teaching strategies are often associated with a deep approach to learning (7, 9, 37, 41) and increased exam performance (16), the cognitive challenge (i.e., LO/HO) and format of the exams is not often reported. In other cases, implementation of active teaching strategies has been met with mixed success (3, 36, 42). Specifically, a lack of structure in the student-centered classroom, particularly for novice learners lacking appropriate schemas to integrate new information with prior knowledge, is a proposed factor to explain the mixed success of these teaching methods (43).

From an assessment perspective, large courses often rely on multiple-choice (MC) questions for ease and efficiency of grading, despite these types of questions frequently being under criticism for primarily requiring LO thinking skills and encouraging rote problem-solving and a surface approach to learning (1, 7, 28,
While a well-designed MC question can assess HO thinking skills in theory (7), the literature has frequently shown that they rarely do in practice (2, 7, 12, 49). In fact, an overwhelming number of Bloom’s level 1 (knowledge) recall-style questions are typical on many undergraduate science assessments (29, 49), promoting memorization of content rather than the comprehension of it (15). Importantly, student anticipation and perception of assessment has considerable influence on their approach to learning in a course (13, 29, 37), including what material they focus on, how much effort they put forth, and how they study (17, 47). Constructed response items (e.g., essay, short and long answer, draw a diagram) are often perceived as more complex than MC assessments, requiring students to demonstrate competence in the skills necessary for academic and work settings (23). Engaging in this complex task involves the creation and construction of information from their knowledge, including example problem solving, graphing, and verbal expression (23). In anticipation of this, students tend to adopt a DA to learning, whereby connections between concepts, application to new scenarios, and understanding of the material are prioritized (14). We teach physiology using an instructor-led lecture in a large classroom setting (~300–400 students). Notably, lecture content is streamlined to focus on integration core principles of physiology (24), with guided problem solving modeled throughout, and assessment is an exclusively written response, requiring students to explain their thoughts as they move through higher levels of Bloom’s Taxonomy. Using a case study approach, we sought to objectively measure whether this teaching and assessment structure could promote a DA to learning and improve HO thinking skills, despite the large class size and instructor-led design commonly associated with a SA to learning and use of LO thinking skills. We predicted that our intentionally designed classroom environment would support meaningful learning, demonstrated by an improvement in performance on assessment questions requiring HO thinking skills over time, as well as a preference for a DA to learning and an increase in students’ DA score over time.

METHODS

All protocols were approved by the University’s Research Ethic Board no. 17–01–004.

Course Context

We recently developed two large (~300–400 student) core physiology courses for health science majors: Human Physiology I (taught by L.A.S.) and Human Physiology II (taught by C.L.M.). Both courses occur over a 12-wk semester (+2-wk final exam period). Human Physiology I (Phys I) is completed by students typically in semester 4 (second year), and Human Physiology II (Phys II) is completed by students typically in semester 5 (third year). Most students will complete both Phys I and Phys II in sequence (e.g., Human Kinetics or Biomedical Science students), although some degree majors only require the completion of Phys I (e.g., Biology students). Students cannot complete Phys II without first completing Phys I. The university is competitive enrollment, with incoming high school averages in the mid-80% and above for the Health Science majors who take these courses (46). The typical composition in these courses is an approximate 70% female to 30% male split (45). Both courses have two 80-min lectures each week in an auditorium-style lecture hall, and while there are different instructors for Phys I and II, they work together closely to ensure consistency in their approach to teaching and assessment of content. There is no laboratory component to these courses.

The main teaching method implemented in both Phys I and II is instructional scaffolding of the core principles in physiology (21, 22, 24). This method refers to both the teaching style (i.e., slower, more supportive lectures to start, building to in-depth integration and problem solving by the end of each unit), as well as the organization of content (early foundational concepts used as building blocks to create progressively bigger systems). Principles are intentionally taught such that students are building a knowledge base and connecting a collection of smaller component ideas as each lecture progresses, demonstrating that each principle is not completely distinct from one another (24). For example, in the first unit of Phys I, students focus on communication principles involved in lifting your foot after stepping on a tack (i.e., membrane transport, bioelectricity, action potential propagation, cell asymmetry, synaptic activity, muscle contraction). In doing so, students effectively “unpack” the smaller component ideas that make up the core principle of information flow (24). With each subsequent unit, these principles are explicitly integrated and expanded on, so that, by the end of the course, students have the biological building blocks necessary to form a complete physiological system. For example, early principles of information flow are used to figure out how the gastrointestinal system works, including slow waves, enteric nervous system signaling, secretion versus absorption, etc. While the class time meetings consist of one instructor guiding the large class at the front of the room, students are provided with a skeleton framework set of notes for each unit of the course and are encouraged to put pen to paper to make note of course content. Students are also encouraged to concurrently work through problem-solving steps with pen and paper following instructor prompts. Instructors begin each unit by providing strong guidance (lecturing) of key concepts. As students become more

1Given consistent trends in demographics each year, these variables were not collected from each student in the present study, rather, these are reflective of institution-wide data.
Each test question is modeled in this way: total 10 marks. Fig. 3. Sample test question from Phys I, approximately week 14.

Knowledgeable, instructors model how to apply core concepts to new situations. By the end of a unit, instructors prompt students with questions to co-construct new systems and solve complex problems as a large group. Therefore, the instructor-guided lecture still remains; however, in these ways, we posit that our delivery and organization of courses content is a modest but manageable departure from a typical PowerPoint, didactic-based lecture.

The primary assessment method utilized in Phys I and II is a long-answer written tests. There are three written tests in each course; Phys I and Phys II each administer two midterms (approximately weeks 5 and 9), and a final exam (exam period, approximately week 14) (Fig. 2). Each test is worth 30, 30, and 40%, respectively, of the final course grade. We acknowledge that students may rely on rote memorization for some concepts in Phys I and II, but contend that long-answer assessments evaluate a student’s ability to explain these concepts and apply them to familiar and new scenarios, rather than an ability to identify correct statements. Therefore, written long-answer tests were deemed most appropriate to challenge and evaluate both students’ LO and HO thinking skills (7, 41). Each midterm is graded out of 30 marks, made up of three 10-mark questions around a central theme or concept from a unit in the course (e.g., action potential propagation). Each question is broken into subquestions, which increase in Bloom’s level (Fig. 3). Questions are intentionally structured to elicit varied levels of thinking, including problem-solving and application, where multiple course concepts must be integrated, applied to new scenarios, and used to make predictions. The final exam is structured similarly, but it is graded out of 40 marks, with four 10-mark questions. Grading is completed by three trained graduate-level teaching assistants under guidance from the instructor. One teaching assistant grades the same test question for all students to improve the consistency of grading, a strategy we have previously shown to be highly accurate and time efficient (32).

Outcome Measures

Ordered thinking skills. Individual test questions were categorized as requiring LO or HO thinking skills, according to the Blooming Biology Tool (BBT) (10), a rubric developed specifically for use in the sciences that aids in the development and categorization of science-related questions among different levels of Bloom’s Taxonomy. Using the BBT as described by Crowe et al. (10), three individuals (faculty, teaching staff, graduate research student), familiar with, but not directly involved in the delivery of, Phys I and II, independently assigned a Bloom’s level to each segment of a long-answer question (i.e., parts a, b, c) on the midterms and finals in Phys I and II. The three individuals then met as a group to discuss any questions that had been assigned different Bloom’s levels, until a clear consensus was reached (33). Questions assigned Bloom’s levels 1 or 2 were aggregated as LO, and those assigned levels 3–6 were aggregated as HO (5, 19). Every assessment contained both LO and HO questions, although the distribution of LO and HO marks varied slightly between assessments. For example, in Phys I, midterm 1 had 15 LO marks and 15 HO marks available, whereas midterm 2 had 17 LO marks and 13 HO marks available. While we acknowledge that a consistent number of LO and HO marks on each test would simplify comparisons between assessments, we felt it important to design assessments to best support student learning and reflect the material covered in lecture, rather than altering the assessment structure for research purposes.

Students’ performance on test questions requiring LO and HO thinking skills was tracked throughout each semester. For each assessment, a LO and HO test grade was calculated for each student, as the percentage of marks they earned on the available LO and HO questions, respectively. For example, if a final exam had 20 LO marks and 20 HO marks available, and a student earned 10 of the LO marks and 12 of the HO marks on the final, the student was given a LO test grade of 50% and a HO test grade of 60%. This calculation was for research purposes only and was not reported to the students during the semester.

Student approach to learning. Student approach to learning was measured at the beginning and end of each course (4 time points total), using the Revised Two-Factor Study Process Questionnaire (R-SPQ-2F) (6) (Fig. 2). The R-SPQ-2F is a 20-item questionnaire commonly used in research to measure students’ learning approaches (9, 19, 34, 40). Ten items on the questionnaire pertain to a student’s DA, and the remaining items pertain to a student’s SA. Deciding which type of question to use on an assessment depends on the goals of the assessment. For example, to evaluate students’ ability to apply core concepts to new situations, we used long-answer questions. For example, a 10-mark question on the final exam: a) Describe the process by which an action potential is propagated down the axon of a myelinated nerve. (4 marks) b) The disease Multiple Sclerosis causes demyelination of previously myelinated nerves. How would this affect the propagation of the action potential? Why? (2 marks) c) How could you change the membrane of the axon in 3b) to restore the action potential propagation in unmyelinated regions (without re-myelinating)? Propose 2 physically sound (but not necessarily medically feasible) changes you could make and explain how they would work. (4 marks)

Fig. 2. Physiology (Phys) I and II each took place over a 12-wk semester (+2-wk exam period). Revised Two-Factor Study Process Questionnaire (R-SPQ-2F) was administered at the beginning and end of Phys I and at the beginning and end of Phys II (4 time points total: R-SPQ-2F-1–4).

Fig. 3. Sample test question from Phys I, midterm 1. Each test question is modeled in this way: total 10 marks, split up into smaller components (e.g., a, b, c), requiring varying levels of ordered thinking skills. HO, higher order; LO, lower order.
10 items pertain to a student’s SA; students respond to each item on a 5-point Likert scale; therefore, they may score a minimum of 10 and a maximum of 50 on each metric. These metrics can further be divided into subscales of motivations (i.e., why students are learning) and strategies (i.e., what students do to learn) such that the 20-item questionnaire has 5 deep motivation questions, 5 deep strategy questions, 5 surface motivation questions, and 5 surface strategy questions. Students may score a minimum of 5 and a maximum of 25 on each subscale metric.

Students were invited to complete the R-SPQ-2F either in class or online, and an incentive to win one of three $50 hospitality gift cards each semester was offered.

Statistical Analysis

Ordered thinking skills. To measure any difference in performance between LO and HO questions on a given assessment in Phys I, paired samples t tests were conducted on each assessment in Phys I. This was then repeated for Phys II. To measure any change in performance on LO or HO test questions across assessments in Phys I, repeated-measures ANOVA was conducted, followed by Bonferroni post hoc testing for between-group significance. This was then repeated across the three assessments in Phys II.

Student approach to learning. To measure preference for DA or SA to learning (and strategies and motives subscales) at a given time point in Phys I, paired samples t tests were conducted at the beginning and at the end of Phys I. This was then repeated for Phys II. To measure any change in DA or SA to learning (and strategies and motives subscales), over the course of Phys I, paired samples t tests were conducted between R-SPQ-2F-1 and R-SPQ-2F-2 in Phys I. This was then repeated for Phys II, by comparing R-SPQ-2F-3 to R-SPQ-2F-4.

All data were analyzed using IBM SPSS Statistics Software version 24. All data are reported as means ± SD. Significance was accepted at P < 0.05.

RESULTS

Ordered Thinking Skills

In Phys I, 349 students completed both midterms and the final exam, and in Phys II 254 students completed both midterms and the final exam.

At midterm 1, students performed significantly better on LO questions compared with HO questions in both Phys I (78.5 ± 20.6 LO vs. 72.9 ± 19.4% HO, P < 0.001) (Fig. 4A) and Phys II (80.5 ± 19.6 LO vs. 69.4 ± 18.4% HO, P < 0.001) (Fig. 4B). However, students’ HO test grade was maintained over the course of the semester in Phys I (72.9 ± 19.4 vs. 74.8 ± 20.7%, P = 0.37) (Fig. 4A) and improved over the course of the semester in Phys II (69.4 ± 18.4 vs. 79.4 ± 14.8%, P < 0.001) (Fig. 4B). By the final exam, students’ HO test grade was significantly better than their LO test grade in both Phys I (74.8 ± 20.7 HO vs. 69.4 ± 17.9% LO, P < 0.001) (Fig. 4A) and Phys II (79.4 ± 24.3 HO vs. 72.2 ± 24.3% LO, P < 0.001) (Fig. 4B).

Ordered thinking data are also displayed in a series of al-luvial diagrams to visualize how student performance on LO and HO test questions changed throughout Phys I and Phys II (Fig. 5). Students’ LO and HO percentage test grades were converted to letter grades using the university’s grading system: 80–100% categorized as “A”; 70–79% categorized as “B”; 60–69% categorized as “C”; 50–59% categorized as “D”; and 0–49% categorized as “F”; and changes in student performance by letter grade are tracked at each time point (test 1, test 2, final exam) in both Phys I (Fig. 5, A and B) and Phys II (Fig. 5, C and D). In both courses, from test 1 to the final exam, there is a decrease in the portion of students earning “A” grades on LO questions [62% test 1 vs. 37% final exam in Phys I (Fig. 5A); 61% test 1 vs. 52% final exam in Phys II (Fig. 5C)] and, therefore, a greater portion of students earning lower grades (i.e., “B”–“F”) on LO test questions. In both courses, from test 1 to the final exam, there is an increase in the portion of students earning “A” grades on LO test questions [45% test 1 vs. 52% final exam in Phys I (Fig. 5B); 35% test 1 vs. 61% final exam in Phys II (Fig. 5D)] and, therefore, a decrease in the portion of students earning lower grades (i.e., “B”–“F”) on HO test questions.

Student Approach to Learning

One hundred and forty-seven students (39% of class) completed both R-SPQ-2F-1 and R-SPQ-2F-2 in Phys I; 76 students (26% of class) completed both R-SPQ-2F-3 and R-SPQ-2F-4 in Phys II.

At all four time points (R-SPQ-2F-1, -2, -3, and -4), students preferred an overall DA to SA (Fig. 6A), and deep motives over...
surface motives (Fig. 6B). There were largely no significant differences in strategy scores, with the exception of R-SPQ-2F-2 (end of Phys I), in which surface strategy was preferred over deep strategy (Fig. 6C).

From the beginning to end of Phys I, and the beginning to end of Phys II, there was no significant change in overall DA or SA to learning scores. Changes in subscale metric scores occurred only in Phys I, not in Phys II, whereby students’ deep motive score increased (14.0 ± 3.5 vs. 14.6 ± 3.5, P = 0.02), deep strategy score decreased (15.0 ± 3.2 vs. 14.2 ± 3.0, P < 0.001), and surface strategy score increased (15.1 ± 3.7 vs. 15.9 ± 3.7, P = 0.002).

DISCUSSION

This study used a case study approach to document student response to two large physiology courses (>300 students), by quantifying academic performance on LO and HO assessment questions, and students’ self-reported approach to learning over the course of both semesters. Although large lecture-based classes are commonly associated with a SA to learning and LO thinking (42, 49), we have shown that intentionally designed lectures grounded in the core principles of physiology (24), which assess students through long-answer constructed response questions, can support a DA to learning and maintain or improve HO thinking. However, we also have shown that LO thinking declines over the course of a semester in this same context.

Intentional Course Design

We describe our primary teaching method as instructional scaffolding, a modest adaptation to the traditional didactic lecture and PowerPoint model. Intentionally, the instructional scaffolding approach implemented by instructors aids students in creating frameworks to organize and integrate new information into their prior knowledge (21), as well as engaging in cognitive processes to support learning (4). For example, the process of elaboration is frequently modeled in Phys, as each lecture loops back to a common problem, creating links to information from previous units (4). Furthermore, the design of the Phys lecture notes encourages students to put pen to paper, compared with the ease of verbatim transcription by keyboard that a laptop encourages (30), therefore allowing students more time to process and engage with new information during lecture. Often in lecture, the instructors will ask students to put all pens down and just listen as a new concept is discussed. Following discussion where information is connected to what students already know or personal experiences, pens are picked up, and the lecture notes for that concept will be completed. By doing so, the instructors intend to model the synthesis, summary, and organization of relevant information following a discussion. Ideally, students are then able to select relevant facts while excluding...
the irrelevant, organizing information in a meaningful way to communicate their understanding on an assessment. This overt modeling of problem solving using core physiological building blocks may be a manageable strategy for other instructors in large classes to consider.

Regarding assessment, we exclusively use long-answer written response questions, where all questions on the six tests spanning Phys I and II are intentionally designed at Bloom’s level ≥ 2, requiring students to integrate ideas from multiple lectures and analyze new situations, rather than testing knowledge of individual facts (41), as is typical with MC tests. Given that this was an uncommon assessment style for most students, we anticipated that student performance and approach to learning would take time to improve after repeated exposure. The strongest HO performance was measured at the end of Phys II (i.e., on the last of 6 tests). This might suggest that a testing effect (i.e., repeated exposure, retrieval practice) contributed to the improvement. However, since questions on subsequent tests were always different, and since LO performance declined across subsequent tests in a given semester, we suggest that any “testing effect” in HO performance is actually support for scaffolding assessments within and across courses, where students are exposed to HO demands early and often. We recognize that large classes are often limited by grading workload, hence reliance on MC questions or fewer number of graded assessments overall, but our findings may motivate instructors to consider incorporating HO questions early in their semester and documenting student progress on these more complex HO skills.

Finally, while our combination of instructional scaffolding and long-answer tests has been shown to support both a DA to learning and HO thinking skills in novice students on an overall (i.e., class average) scale, we acknowledge that our alluvial diagrams show that some students did not show this desired effect over time. Therefore, the significant improvement in HO performance in Phys II may be particularly impressive. Our study is strengthened and adds important information to the literature by considering student performance longitudinally by LO and HO skills, as these findings would remain undiscovered if documented only at a single point.


depth

**Fig. 6.** Values are means ± SD. Revised Two-Factor Study Process Questionnaire (R-SPQ-2F) scores at the beginning and end of Physiology (Phys) I (R-SPQ-2F-1 and R-SPQ-2F-2, n = 147), and the beginning and end of Phys II (R-SPQ-2F-3 and R-SPQ-2F-4, n = 76) are shown. A: student overall approach to learning score. B: student motive subscale score. C: student strategy subscale score. *Paired samples t test revealed significant differences between deep and surface learning of a given metric at each time point, $P < 0.05$.

**Ordered Thinking Skills**

In support of our hypothesis, HO performance improved over time in Phys II, although there was no significant change in HO performance from beginning to end of Phys I. However, when these HO results are interpreted alongside the concomitant decline in LO performance within each semester, the initial perceived lack of improvement in HO performance in Phys I may instead be considered a maintenance of performance. Therefore, the significant improvement in HO performance in Phys II may be particularly impressive. Our study is strengthened and adds important information to the literature by considering student performance longitudinally by LO and HO skills, as these findings would remain undiscovered if documented only at a single point.

We were surprised to find that students’ LO test grade decreased over the course of a semester, particularly since this pattern was consistent in both Phys I and Phys II. Bloom’s Taxonomy suggests that, to build HO thinking skills where information is truly understood, each thinking level must be followed in steps up the hierarchy (37). Casagrand and Semsar (8) have reported improved performance on both LO and HO exam questions following 4 yr of incremental changes to course
instruction and assessment design in a senior neurophysiology course. Willingham (48) argued that students who focus their efforts on practicing facts until they are memorized can more easily apply this knowledge to HO learning, decreasing cognitive load (21). Within our Human Physiology course, however, the retrieval and reproduction of facts is not emphasized, rather big picture core principles are the focus within the lecture. As such, students may focus more on connecting and integrating concepts as each semester continues, potentially glossing over the specific details required to earn full marks on LO questions, without completely evading the understanding of course content either (39). Additionally, the constructed response style of our tests means that LO questions are not “all or nothing,” as is typical with MC tests, such that a LO score of 70% should not be rigidly interpreted as getting 30% of concepts incorrect. Rather, students typically earn part marks on LO questions by being able to explain some key elements of a concept, but may not articulate their responses as precisely or thoroughly to earn full marks. It could be possible that a threshold level of LO understanding is all that is necessary to be able to apply core concepts to novel situations and predict outcomes (thus being successful on HO questions). Notably, even as LO performance declined, the class average on these questions still remained in the very high 60% to low 70% range. Therefore, it is important to not overinterpret this finding of a decrease of a LO performance as an indication that students know less information as the semester goes on, rather that there is a shift toward supporting HO thinking.

A final important consideration in this study is the change in performance over time. A student’s cognitive load may be increasingly at capacity by the end of each semester as concepts have been building throughout the Human Physiology course, compounded with studying information from other courses students are taking concurrently (typically 4 others per semester), especially during the final exam period. Therefore, students may have increased difficulty in retaining and explaining the specifics required to perform well on LO questions. Considering that LO performance consistently worsens, but this pattern does not occur in HO performance, it appears that performance of HO thinking skills may be protected from the end-of-semester cognitive overload. This suggests that student grades may actually improve if instructors were to emphasize HO questions on final exams.

Student Approach to Learning

At all four time points measured (R-SPQ-2F-1−4), students reported a preference for a DA to learning over a SA to learning, deep motives over surface motives, but no consistent preference for deep or surface strategies. Combined, these findings suggest that, at the very least, students can maintain a DA to learning over the duration of two large, lecture-based courses; however, they do not appear to adopt a deeper approach over time after exposure to our course context. The DA scores recorded in our study (29.0, 28.8, 31.6, and 30.8, respectively) are roughly comparable to those presented in the literature, although most measures of student approach to learning in the literature are recorded at a single time point (18, 38, 40), and the level of detail regarding the context in which the learning takes place is typically limited. For example, a DA score of 33.4 was reported for fourth-year Faculty of Science students (n = 30) from a variety of programs (e.g., Biology, Computer Science, Physics) (40), but no information about specific course contexts was provided. Further challenges with interpreting available data include the absence of a guideline or standard to determine what constitutes a “high” or “low” approach to learning score, and the lack of clarity regarding what magnitude of change in approach to learning score would be necessary to translate into a meaningful practical change for students. Finally, while studies often report on overall DA, the relative contribution of subscales (i.e., motives vs. strategies) that make up the overall score are not often reported. Therefore, our study expands on the existing literature by reporting a series of approaches to learning scores pre- and postexposure to an intentionally designed learning environment, providing details on the course context where these values were obtained, and reports both overall approach to learning values and subscale metrics. As such, this study may serve as a valuable reference point to researchers attempting to contextualize their own student approach to learning data.

We acknowledge that interpretation of these findings must also consider the potential for student response bias, even though all responses were anonymized and did not contribute toward grades within the class. Students were asked to complete R-SPQ-2F-1 (baseline approach to learning) while thinking about how they approached learning in Introductory Biochemistry, a prerequisite for Phys I that uses lecture and MC assessment formats commonly shown to reinforce a SA (1, 7, 41). The unexpected reporting of a preference for a DA on this baseline survey may be influenced by social desirability, wherein students respond to the questionnaire such that they appear to be “good” students at the beginning of the term, even if they do not accurately reflect their experiences (25). Additionally, survey respondents (response rate: 39% in Phys I and 26% in Phys II) were on average academically stronger than the full class (+5%), although their final grades did span the spectrum from “F” to “A”. As such, the consistent finding of an overall preference for a DA over a SA to learning may be driven by students’ responding in a socially desirable way, or academic differences in survey respondents. However, of note, this subset of students still showed the same patterns in HO and LO performance over the duration of each course as the full cohort, albeit at a slightly higher absolute grade. Future research may consider collecting further information from students to control for these potential differences in student groups (e.g., incoming GPA, overall GPA).

Conclusions

Using a case study approach, we documented that a large physiology course sequence can support a DA to learning and improve students’ HO thinking skills related to core physiology principles over time. These findings may encourage other educators to critically review their teaching and assessment strategies and consider incremental changes, including distinguishing between LO and HO student performance to ensure that student outcomes align with the overall learning outcomes of their course. Awareness of students’ motivations and strategies in their approach to learning may also influence this relationship, but the specific impact would require further consideration within a given course context. Overall, our results support that large enrollment courses with instructor-led lectures and test-based assessment do not preclude students from achieving HO thinking skills that are required for physiology. Future research could consider a more
experimental approach to differentiate between the specific role of individual factors.

GRANTS
This research was supported by Social Sciences and Humanities Research Council of Canada Doctoral Fellowship Award.

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

AUTHOR CONTRIBUTIONS
J.O.H., C.L.M., and K.L.R. conceived and designed research; J.O.H. and K.L.R. performed experiments; J.O.H., C.L.M., L.S., J.M.T., and K.L.R. analyzed data; J.O.H., C.L.M., and K.L.R. interpreted results of experiments; J.O.H., C.L.M., L.S., J.M.T., and K.L.R. edited and revised manuscript; J.O.H., C.L.M., L.S., J.M.T., and K.L.R. approved final version of manuscript.

REFERENCES
1. Alonso M, Stella C, Galagovsky L. Student assessment in large-enrollment biology classes. Biochem Mol Biol Educ 36: 16–21, 2008. doi:10.1002/bme.20147
2. Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, Raths J, Wittrock MC. A Taxonomy for Learning, Teaching, and Assessing; A Revision of Bloom’s Taxonomy of Educational Objectives. New York: Pearson, 2000.
3. Andrews TM, Leonard MJ, Colgrove CA, Kalinowski ST. Active learning not associated with student learning in a random sample of college biology courses. CBE Life Sci Educ 10: 394–405, 2011. doi:10.1187/cbe.11-07-0061.
4. Arnold KM, Umanath S, Thio K, Keilty WB, McDaniel MA, Marsh EJ. Understanding the cognitive processes involved in writing to learn. J Exp Psychol Appl 23: 115–127, 2017. doi:10.1037/xpa0000119.
5. Bibler Zaidi NL, Grob KL, Yang J, Santen SA, Monrad SU, Miller JM, Purkiss JA. Theory, process, and validation evidence for a staff-driven medical education exam quality improvement process. Med Sci Educ 26: 331–336, 2016. doi:10.1007/s40670-016-0275-2.
6. Biggs J, Kember D, Leung DYP. The revised two-factor Study Process Questionnaire: R-SPQ-2F. Br J Educ Psychol 71: 133–149, 2001. doi:10.1348/000709901158413.
7. Biggs JB, Tang CS. Teaching for Quality Learning at University: What the Student Does (4th ed.). New York: McGraw-Hill Education, 2011.
8. Casagrand J, Sensar K. Redesigning a course to help students achieve higher-order cognitive thinking skills: from goals and mechanics to student outcomes. Adv Physiol Educ 41: 194–202, 2017. doi:10.1152/advan.00070.2016.
9. Christensen Hughes J, Mighty J. Taking Stock: Research on Teaching and Learning in Higher Education. Kingston, ON, Canada: McGill-Queen’s University Press, 2010.
10. Crowe A, Dirks C, Wenderoth MP. Biology in bloom: implementing Bloom’s Taxonomy to enhance student learning in biology. CBE Life Sci Educ 7: 368–381, 2008. doi:10.1187/cbe.08-05-0024.
11. Cuseo J. The empirical case against large class size: adverse effects on the teaching, learning, and retention of first-year students. J Fac Dev 21: 5–21, 2007.
12. Driessen E, Van Der Vleuten C. Matching student assessment to problem-based learning: lessons from experience in a law faculty. Stud Contin Educ 22: 235–248, 2000. doi:10.1080/0713695731.
13. Entwistle N, Tait H. Approaches to studying and perceptions of the learning environment across disciplines. New Dir Teach Learn 1995: 93–103, 1995. doi:10.3124/rdl219956413.
14. Entwistle NJ, Entwistle A. Contrasting forms of understanding for degree examinations: the student experience and its implications. High Educ 22: 205–227, 1991. doi:10.1023/B:BF00132288.
15. Flores MA, Veiga Simao AM, Barros A, Pereira D. Perceptions of effectiveness, fairness and feedback of assessment methods: a study in higher education. Stud High Educ 40: 1523–1534, 2015. doi:10.1080/03075079.2014.881348.
16. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci USA 111: 8410–8415, 2014. doi:10.1073/pnas.1319030111.
17. Gibbs G, Simpson C. Conditions under which assessment supports students’ learning. Learn Teach High Educ 1: 3–31, 2005.
18. Gijbels D, Segers M, Struyf E. Constructivist learning environments and the (im)possibility to change students’ perceptions of assessment demands and approaches to learning. Instr Sci 36: 431–443, 2008. doi:10.1007/s11251-008-9064-7.
19. Jensen JL, McDaniel MA, Woodward SM, Kummer TA. Teaching to the test...or testing to teach: exams requiring higher order thinking skills encourage greater conceptual understanding. Educ Psychol Rev 26: 307–329, 2014. doi:10.1007/s10648-013-9248-9.
20. Kerr A. Teaching and Learning in Large Classes at Ontario Universities: An Exploratory Study. Toronto, ON, Canada: Higher Education Quality Council of Ontario, 2012.
21. Kirschner PA, Sweller J, Clark RE. Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educ Psychol 41: 75–86, 2006. doi:10.1080/014434105005809.x.
22. Larkin M. Using Scaffolded Instruction To Optimize Learning (Online), ERIC Clearinghouse on Disabilities and Gifted Education, Council for Exceptional Children. https://eric.ed.gov/?id=ED474301 [8 April 2020].
23. Lukshele R, Thissen D, Wainer H. On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. J Educ Meas 31: 234–250, 1994. doi:10.1111/j.1745-9849.1994.ber.2185.x.
24. Michael J, Modell H, McFarland J, Cliff W. The “core principles” of physiology: what should students understand? Adv Physiol Educ 33: 10–16, 2009. doi:10.1152/advan.00139.2008.
25. Miller AL. Investigating Social Desirability Bias in Student Self-Report Surveys. Annual Forum of the Association for Institutional Research. Toronto, Ontario, May 21–25, 2011, p. 1–18.
26. Miller CJ, McNear J, Metz MJ. A comparison of traditional and engaging lecture methods in a large, professional-level course. Adv Physiol Educ 37: 347–355, 2013. doi:10.1152/advan.00050.2013.
27. Modell HL. Helping students make sense of physiological mechanisms: the “view from the inside”. Adv Physiol Educ 31: 186–192, 2007. doi:10.1152/advan.00797.2006.
28. Momens J, Offerdahl E, Kryjevskia M, Montplaisir L, Anderson E, Grosz N. Using assessments to investigate and compare the nature of learning in undergraduate science courses. CBE Life Sci Educ 12: 239–249, 2013. doi:10.1187/cbe.12-08-0130.
29. Momens JL, Long TM, Wyse SA, Ebert-May D. Just the facts? Introducing undergraduate biology courses focus on low-level cognitive skills. CBE Life Sci Educ 9: 435–440, 2010. doi:10.1187/cbe.10-01-0001.
30. Mueller PA, Oppenheimer DM. The pen is mightier than the keyboard: advantages of handover of laptop note taking. Psychol Sci 25: 1159–1168, 2014 [Erratum in Psychol Sci 29: 1565–1568, 2018]. doi:10.1177/0956797614524581.
31. Mulryan-Kyne C. Teaching large classes at college and university level: challenges and opportunities. Teach High Educ 15: 175–185, 2010. doi:10.1080/13562511003620001.
32. Murrant CL. Creative teaching assistant organization to maintain an Integrative Physiology course with 440 students. Adv Physiol Educ 31: 180–185, 2007. doi:10.1152/advan.00061.2006.
33. Murrant CL, Dyck DJ, Kirkland JB, Newton GS, Ritchie KL, Tishinsky JM, Bettger WJ, Richardson NS. A large, first-year, introductory, multi-sectional biological concepts of health course designed to develop skills and enhance deeper learning. Can J High Educ 45: 42–62, 2015.
34. Newton G, Martin E. Research and Teaching: Blooming, SOLO Taxonomy, and Phenomenography as Assessment Strategies in Undergraduate Science Education. J Coll Sci Teach 043: 78–90, 2013. doi:10.2505/j4csc103_043_02_78.
35. OCUFA. Data Check: Class Sizes Continue to Grow at Ontario’s Universities (Online). https://ocufa.on.ca/blog-posts/data-check-class-sizes-continue-to-grow-at-ontarios-universities/12 July 2018.
36. Prince M. Does active learning work? A review of the research. J Eng Educ 93: 273–281, 2004. doi:10.1002/0471787634.bmb.20147.
37. Pungerente MD, Badger RA. Teaching Introductory Organic Chemistry: “blooming” beyond a simple taxonomy. J Chem Educ 80: 779, 2003. doi:10.1021/cf0807797.
38. Rajaratnam N, D’cruz SM, Chandrasekhar M. Correlation between the learning approaches of first year medical students and their performance in multiple choice questions in physiology. *Natl J Integr Res Med* 4: 43–48, 2013.

39. Ramsden P. *Learning to Teach in Higher Education* (2nd ed.). Oxon, UK: Routledge Falmer, 2003.

40. Sabourin B. Identifying Student Approaches to Learning: Undergraduate Student Perceptions of Teaching and Learning at the University of Windsor. Windsor, ON, Canada: University of Windsor, 2016.

41. Scouller K. The influence of assessment method on students’ learning approaches: multiple choice question examination versus assignment essay. *High Educ* 35: 453–472, 1998. doi:10.1023/A:1003196224280.

42. Silverthorn DU, Thorn PM, Svinicki MD. It’s difficult to change the way we teach: lessons from the Integrative Themes in Physiology curriculum module project. *Adv Physiol Educ* 30: 204–214, 2006. doi:10.1152/advan.00064.2006.

43. Struyven K, Dochy F, Janssens S, Gielen S. On the dynamics of students’ approaches to learning: The effects of the teaching/learning environment. *Learn Instr* 16: 279–294, 2006. doi:10.1016/j.learninstruc.2006.07.001.

44. Tobias S. *Revitalizing Undergraduate Science: Why Some Things Work and Most Don’t*. Tucson, AZ: Research Corp., 1992.

45. University of Guelph. *Fact Book 2017–2018* (Online). https://www.uoguelph.ca/iar/fact-book [16 Dec 2019].

46. University of Guelph. *Specific Subject Requirements—Ontario* (Online). https://admission.uoguelph.ca/ontariosubreqs [16 Dec 2019].

47. Villarroel V, Boud D, Bloxham S, Bruna D, Bruna C. Using principles of authentic assessment to redesign written examinations and tests. *Innov Educ Teach Int* 57: 38–49, 2020. doi:10.1080/14703297.2018.1564882.

48. Willingham D. *Why Don’t Students Like School? A Cognitive Scientist Answers Questions About How the Mind Works and What It Means for the Classroom*. San Francisco, CA: Jossey-Bass, 2009.

49. Wood WB. Innovations in teaching undergraduate biology and why we need them. *Annu Rev Cell Dev Biol* 25: 93–112, 2009. doi:10.1146/annurev.cellbio.24.110707.175306.