Michael deBraga, UNIVERSITY OF TORONTO MISSISSAUGA,
michael.debraga@utoronto.ca
Cleo Boyd, UNIVERSITY OF TORONTO MISSISSAUGA, cleo.boyd@utoronto.ca
Shahad Abdulnour, UNIVERSITY OF TORONTO MISSISSAUGA,
shahad.abdulnour@utoronto.ca

Using the Principles of SoTL to Redesign an Advanced Evolutionary Biology Course

ABSTRACT

A primary goal of university instruction is the students' demonstration of improved, highly developed critical thinking (CT) skills. However, how do faculty encourage CT and its potential concomitant increase in student workload without negatively impacting student perceptions of the course? In this investigation, an advanced biology course is evaluated after structural changes (implemented in 2010) met with a poor student evaluation of the course and the instructor. This analysis first examines the steps used to transform a course to encourage CT and then explains how it can be assessed. To accomplish these goals, the instructor collaborated with an educational developer to redesign the course using a philosophy informed by SoTL. This approach, as we see it, represents a set of principles that demand transparency in the development and application of strategies whose aim is to encourage student learning. However, the SoTL approach would be insufficient to simply promote a set of strategies without some mechanism for evaluating its efficacy. Therefore, we designed a "Graded Response" (GR) multiple-choice test to measure CT development and hence to properly evaluate whether the strategies embedded in our SoTL-informed course redesign have adequately met our goals.

KEYWORDS

course redesign, constructive alignment, critical thinking, reflected practice

INTRODUCTION

The Scholarship of Teaching and Learning (SoTL) begins with moments of reflection on what students are learning (or not), and why. Brookfield (1997) emphasizes the role of reflection whereby instructors should examine their performance through a lens of what he termed "critical mirrors" (Brookfield, 1997, p. 19). He encourages inviting criticism from peers to facilitate deeper thinking about teaching and learning, a necessary part of course redesign: "When our peers listen to our stories and then reflect back to us what they see and hear in those stories, we are often presented with a version of ourselves and our actions that comes as a surprise" (Brookfield, 1997, p.19). Ghaye (2011) also argued for reflection, what he termed "Reflected Practice," as a key attribute in the development of a course aimed at encouraging deeper understanding. Reflection also plays a role in

15

Biggs's (1996 & 1999) notion of "constructive alignment" in course design: only through personal reflection on the part of the instructor could the alignment of course objectives and assessment tasks be integrated in a manner that encourages deeper learning. Brookfield also connects reflection on the teaching and learning context with the goals for student when he states, "Accurate assessment springs from an informed understanding of the phenomenon being studied, and the assessment of critical thinking is no exception" (Brookfield, 1997, p. 19). In summary, reflection should play a key role in course redesign aimed at critical thinking (CT), as well as the methods for assessing the revisions.

Although reflection isn't always listed explicitly as a step in SoTL, it's implied in many of its definitions and descriptions. Felten (2013) eloquently details five principles for evaluating SoTL: inquiry into student learning, grounded in context, methodologically sound, conducted in partnership with students, and appropriately public (p. 122). His is a useful list in developing SoTL projects and programs, but I argue that these "principles of good practice in SoTL" are also principles of good practice in course redesign aimed at deeper understanding. His principles coupled explicitly with reflected practice (Ghaye, 2011) and constructive alignment methodologies (Biggs, 1996; Biggs & Tang, 2007) are here presented as a template for course design. Furthermore, to test the effectiveness of the proposed template, this essay will assess student learning in the course redesign using a Graded Response (GR) test, which is based on the Watson Glaser Critical Thinking Appraisal® (WGCTA – Watson & Glaser, 2006).

Altogether, this framework establishes a SoTL-guided approach to course development and redesign grounded in reflection, constructive alignment, an exploration of the pertinent literature, and experimental analysis. This paper is divided into two parts: the first part will explain the rationale for applying the principles of SoTL, whereas the second part provides the study of an example of a course resulting from this redesign framework. This course underwent a number of changes aimed at increasing the critical thinking in an upper-year biology course in a large Canadian university with a reputation for academic research excellence. The cohort under investigation represents a highly diverse demographic comprised of domestic and international students.

PART 1: APPLYING FELTEN'S SOTL PRINCIPLES TO COURSE REDESIGN

"Inquiry focused on student learning"

Applying Felten's first principle of SoTL to course redesign methodology ensures that the instructor starts by reflecting on the students' learning. It encourages a specificity by situating that learning (or lack thereof) in the context of the course's precise structure (i.e., lecture-delivery model, assignment goals and expectations, alignment with specific outcomes, and assessments). It is essential—in SoTL as in course redesign—to begin by asking what it is that we want our students to learn and perhaps even more importantly how we want the students to learn it. In fact, Trigwell (2013) claims that students' approach to learning—surface or deep—reflects the presence (deep) or absence (surface) of the instructor's SoTL approach to teaching.

There is ample evidence to suggest that today's student functions in a learning environment that is based on the avoidance of failure (Covington, 1984; Keeley, Shemberg, Cowell, & Zinnbauer, 1995; Seifert, 2004). Hounsell, McCune, Litjens, & Hounsell

(2005) observed that early-year students are much more likely to exhibit a pronounced fear of failure, while upper-year students have a fairly sound comprehension of what is expected within their field. However, this recognition of what is expected may not be manifested in the latter students' attitude toward a particular course, as many university students resist active engagement and hence remain passive in their approach to learning, which students perceive as carrying less risk (Haas & Kelley, 1998; Leshowitz, DiCerbo, & Symington 1999; Seifert, 2004). Indeed, even students who already possess many of the relevant CT skills may resist exercising such skills in the absence of direction from an "esteemed peer" (to use Mathews' and Lowe's term [2011]). This failure-aversive resistance to CT has been discussed by Keeley et al. (1995), Tsui (2002), and Mathews and Lowe (2011), who argue that engaging in CT requires a supportive environment. Similarly, Leshowitz, DiCerbo, & Symington (1999) claim that "pseudoknowing" or surface learning can be replaced by "reasoned knowing" only if the appropriate support mechanisms are in place. Starting with this concern about student learning, course redesign frames the course with guidance, support, and opportunities for students to successfully demonstrate this learning.

"Grounded in context"

Felten claims that good practice in SoTL is also grounded in its specific contexts. Evidence suggests that the most effective learning also takes place when it is context-based (Entwistle, 2000). This principle is critical in order to fully apply interventions that encourage the desired learning—especially CT. For instance, performance must be modeled for the students in a manner that addresses both the prior iterations the course, as well as the standards of the discipline. Simply demonstrating the preferred method without these contexts is insufficient. Actively deconstructing a sample paper with the class within the context of disciplinary and course expectations is another approach. In other words, instructors model their own approach to evaluating a paper by examining the questions that are proposed by the authors and then consider the author's findings in order to evaluate whether the questions have been addressed in a robust manner. Key components would include assessing how effectively questions are addressed, examining competing methodologies, and deriving a summary of the differences and similarities. By building into a course plenty of exercises or tutorials that encourage the investigative process, students learn how to recognize the quality of arguments and the difference between an opinion and a supported hypothesis or claim. Situating an intervention within the larger context of the learning goals of the course and the discipline helps scaffold the support at a time when the students most need to practice the relevant skill.

"Methodologically sound"

Another key attribute of Felten's (2013) principles for evaluating SoTL are the strength or soundness of its methodology, again a criterion that can be applied to course redesign. The previous steps in this course redesign framework require instructors to revisit the goals and contexts for student learning. Trigwell (2013) encourages us to start earlier by evaluating our approach to teaching. Assessments, for example, should reflect both the learning goals and the intentions in teaching. Trigwell describes assessment intended to measure content acquisition as an "Information Transfer / Teacher-Focused Model" (ITTF), a conception of teaching that limits students' ability to experience deeper un-

derstanding (p. 98). Revisiting the approaches to teaching and the learning provides a sounder course redesign methodology, one that is more likely to lead to what Trigwell refers to as a "Conceptual Change / Student-Focused" model (CCSF).

"Conducted in partnership with students"

Felten (2013) describes a principle of mutually supportive roles played by both instructor and student in SoTL, a partnership that also works well in course redesign. Scott, Lisagor, and Marachi (2009) describe the changing face of American universities as involving a paradigm shift from a pedagogy that focused on direct instruction to one that construes the entire instructor-student experience in terms of collaboration. In essence, this philosophy points to a responsibility on the part of the instructor to provide suitable incentives that encourage students to participate in their own learning process. The literature describes many learning environments as a by-product of the contemporary university cohort (Abualrub, Karseth, & Stensaker, 2013; Aldridge & Fraser, 2000; Aldridge, Dorman & Fraser, 2004; Blair, 2013) The absence of a SoTL approach to course design may directly contribute to the course's failure, as the instructor has misjudged the students' learning and various contexts and leaves unexamined the entire approach to the classroom. This can leave students ill prepared and not properly guided through the investigative and reflective behaviours that are vital for creating a CCSF environment more likely to support CT. Lizzio, Wilson, & Simons (2002) point out that students appear to develop a better and, hence, deeper understanding of course material if they have more time to reflect on and reinforce their learning. As a result, courses that are perceived as having a very high workload and affording less time for reflection are, in Lizzio et al.'s (2002) estimation, not as likely to promote CT.

"Appropriately public"

Felten's final principle of SoTL supports the commitment of instructors to the sharing of their SoTL. The reasons for going public in SoTL apply to course redesign as well. Making available for peer review the process and details of course redesign provides a means for other instructors to examine, evaluate, and potentially adapt to their own courses. This SoTL-inspired template thus serves as a way to make course redesign public as a transparent and thus reproducible framework. A key trait of this course redesign template, therefore, is not only to provide a method for redesigning a course, but also to ensure that the entire process can be assessed through its exposure to academic scrutiny.

PART 2: SOTL REDESIGN IN PRACTICE: AN APPLICATION

The Redesign

As a sabbatical replacement in 1996, deBraga taught Major Features of Vertebrate Evolution (BIO356), an advanced evolutionary biology course. The course had been delivered in a traditional manner with a focus on the anatomical diversity of vertebrate life throughout the earth's history, and its variety of instructors focused on measuring content acquisition and, to a much lesser degree, on assessing students' metacognitive abilities and "deeper understanding." In all iterations of the course, its structure remained relatively static until 2010.

In the 2010 iteration of BIO356, the laboratory components (including fossil examination labs) were replaced with tutorials designed to support new assignments aimed at

offering students extra support for their learning. However, it failed: the absence of the opportunities for reflection and feedback on the part of the students during the tutorials proved problematic, and the changes in the course had an unanticipated and significantly negative impact vis-à-vis the learning centered ideal embedded in Trigwell's CCSF approach (Brookfield, 1997; Froyd & Simpson, 2010). The dramatic change in focus from traditional, predominantly content-driven assessment resembling the ITTF model to a more theoretical set of investigative assignments (i.e., assignments that tasked students to evaluate the merits of competing hypotheses) without a concomitant consideration of how the students would experience and learn (or not) these new approaches proved problematic. Student opinion surveys (SOS) demonstrated the negative consequences of the changes (see Table 2), particularly suggesting a course design lacking in Felten's first principle of using the student learning as the touchstone.

In addition, little attention had been paid to the demographic nature of the course's 2010 version and how much had changed since my previous experience over a decade earlier (1996). An entirely new generation of students had arrived with an institutional memory of a very specific view of the course and its structure. Students were not well prepared for the complexity of the tasks that these new assignments demanded, and the course revision hadn't started with their learning in mind, considered the implications of the context of this specific student body, or offered appropriate support mechanisms, so an unanticipated but perhaps predictable resistance to the new course structure was observed (Beachford & Beachford, 2010; Kyndt, Dochy, Struyven, & Cascallar, 2011; Taylor, 2006). Finally, the approach to assessment was in need of review. As we began the course redesign, we paid particular attention to how the assessments were designed (or not) for the specific student learning goals of CT.

Beginning with the student learning

The active role of the educational developer at the start of the process and in on-going assessment of the changes in the course strengthened the redesign methodology and ensured proper scrutiny (Felten's 3rd principle). This collaboration also ensured that we started with—and continued to focus on—the student learning. The entire course syllabus was revised by establishing the learning outcomes of the course and using a backward design to accomplish those goals (Prideaux, 2003), an approach that meets Felten's first principle. That process eventually led to a more interactive approach to delivery of in-class materials, including CT-focused tutorials and assignments. The resulting GR test represented an entirely new form of multiple choice(MC) test, which was based on the inference capacity of the WGCTA. More significantly, it changed the primary mode of learning and assessment in the course from a surface ITTF to one which demanded a deeper understanding on the part of the students.

A Focus Group Approach to "Partnership with Students"

The implementation of a focus group or some other intervention for gathering real-time feedback from the students established what Felten might consider more of a partnership with the students in supporting their learning. The focus group was used as a qualitative assessment of student feedback during the 2011 and 2012 iterations of the course. This feedback, coupled with student opinion survey (SOS) data, helped to frame and reinforce the general student perceptions of the course. The use of a focus group permits

a real time reflective component as well as a post-course reflective component, both of which are aimed at making sure the instructor understands the ever-changing student demographic while adhering to intended learning goals.

Ultimately among the different focus group types, it was decided to pursue an exploratory methodology (Morrison, 1997), which benefits from the informal nature of its design. The premise of using this methodology is the sense of "safety" that the students feel with the educational developer when discussing their attitude toward the course and the instructor. The focus group session took place during a regularly scheduled tutorial session when neither the course instructor nor the TA was present. Students were assured of the anonymity of their responses and were encouraged to be honest in their evaluation of the course and the instructor. The questions that were put to the students were divided into two categories: on the one hand, lecture and tutorial content and, on the other, evaluation of the instructor. (See Appendix 1 for focus group questions.) The questionnaire served as an opportunity for the students to consider their overview of the course while also providing feedback to the instructor at a point that would permit some real-time reflection. The value was qualitative, and it was not our intent to formally review the responses. Hence, responses were blended, and the educational developer arranged a time to discuss the feedback with the instructor. The general impression was one of support for the instructor and notably for the helpful nature of the tutorials, but students on the whole were concerned that specific direction was lacking. More precisely, students were afraid that the question-generating nature of the lecture material and assignment structure precluded clear parameters, especially concerning what they needed to know, echoing what much of the literature suggests about student fears of taking risks and desiring more certainty and security (Haas & Kelley, 1998; Leshowitz, et al., 1999; Seifert, 2004).

The Intervention

In the syllabus, in the first class, and throughout the course, the instructor was very forthcoming that the instructional paradigm for this iteration of the course was going to be much more demanding than previous iterations. Ghaye (2011) considers the time and effort that are required for this level of interaction as needing justification. As this course transformed from its 2010 iteration through into the final 2013 iteration, it became clear that modeling the appropriate student behaviors was necessary, especially in terms of reflective practice as a foundation for CT skills.

One tutorial in particular was specifically designed to prepare the students for an assignment that tasked them with evaluating the merits of competing phylogenetic hypotheses. In this instance, the students would have to assess the degree of accuracy of the competing arguments or evaluate the degree of "truth" of the competing views. The idea of "truth" or truthfulness is an extension of the underlying philosophy of the WGCTA and is fundamental to the GR test, the specific assessment we developed during this course redesign. The most critical aspect of the tutorial design is not summative but instead a formative intervention.

Students were paired up randomly or, if numbers necessitated, put into groups with a maximum of three participants. Students were then each assigned a different article that proposed competing hypotheses and were asked to identify inconsistencies or differences in both the arguments and methodologies. The course instructor or TA then

asked each group to discuss their paper and determine if a consensus could be reached regarding the proposed hypotheses. After a short period of time, groups were asked to deliver their criticism, which was then discussed within the class. The value to this approach is that students were free to present their opinions in a risk-free, low-stakes environment. All student opinions, which were supported with evidence, were assumed to have value as they were part of the reflective practice that had been emphasized at the start of the tutorial. The aim of the exercise was to help in the development and honing of formal arguments that could be used to better support a particular claim. Although not all groups were able to present their results during the one-hour tutorial, there were opportunities to continue discussions of the activity (and the overall course) through a discussion board. Students were encouraged to participate but were not given credit for doing so. Instead, either the course instructor or TA regularly addressed the concerns on the discussion board that would benefit from looking to the course objectives and emphasized the importance of active participation in these discussions for their assignments and other summative assessments that were to follow.

Assessing the "soundness" of the redesign methodology

To assess the interventions implemented in the course redesign between 2010 and the fully supported courses of 2012 and 2013, it was necessary to quantitatively measure the impact of the modified tutorials and the GR test. This process focused on the efficacy of the WGCTA as the template for the course's new assessment (the GR test), as well as data collected from the various course iterations (pre-intervention vs. post-intervention) regarding any impact on CT. Additionally, this process involved comparing SOSs in the pre- and post-intervention iterations of the course to identify the challenges resulting from such a significant structural change in the course, notably transitioning from measuring content acquisition to an investigative design.

The WGCTA (Watson & Glaser, 2006) was designed to test CT skills among professionals and students alike. The version of the test applied in this investigation represents the WGCTA "Short Form" which was first published in 1994 and has been revised as recently as 2005. The test is designed to measure five specific skills: inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments. The WGCTA has been used quite extensively in the professional world and notably in professional university programs such as nursing, where it has repeatedly demonstrated its reliability as a measure of CT (+/- 1.96 at 95% confidence limits; Magnussen, Ishida, & Itano, 2000). Of the five WGCTA skills, the inference testing demonstrated the greatest application to our GR test design, based on how the logic inherent in this type of question[1] suits the aim of the GR test and serves as a means to encourage CT. Notably, the form of this question requires that a student can determine the degree of truthfulness for each of the statements. In the example below, we compare a question from a traditional pre-intervention MC test (1a) to the modified GR test questions (1b) from the course redesign. Note that the bolded response in 1a refers to the correct response. In figure 1b a full mark is awarded only if the correct order from best to worst (i.e., 1= best, 4 = worst) is assigned by the student. If the best and the worst responses are identified correctly but the remaining order is incorrect, only one-half mark is awarded. Alternatively, if only the best choice is selected and the rest are incorrect, one-quarter mark is awarded. A grade of zero is applied only if all options are incorrectly ordered.

1a. (WINTER 2010) In the view of the fossorial origin of snakes, . . .

   A. mosasaurs are believed to be the closest sister-group because of their long tails.

   B. mosasaurs are believed to be the closest sister-group because of their streptostylic skulls.

   C. amphisbaenians are believed to be the closest sister-group because of their burrowing habits.

   D. their delicate skulls are interpreted as being an adaptation for burrowing.

   E. legs are believed to have served a burrowing role in early members.

1b. (WINTER 2012-13) Which statement best reflects the current understanding? (Rank the statements.)

   4-A. Snakes have separate, independently moveable lower jaw elements.

   2-B. Snakes possess a ligamentously attached jaw symphysis permitting extensive flexibility during feeding.

   1-C. Snakes possess a lower jaw symphysis that is loosely connected by a ligament permitting autonomous movement of each half of the lower jaw thereby facilitating the ingestion of large prey.

   3-D. Snakes have a flexible lower jaw designed to permit the ingestion of large prey.

In question 1b, the best response (1-C) corresponds to the "TRUE" statement whereas the worst response (4-A) would correspond to the "FALSE" statement as expressed in the WGCTA test. Therefore, students were no longer required to simply identify correct responses but were, in essence, being tasked with evaluating why some responses are better than others. This requirement forces the student to justify a rationale with respect to how they rank the statements and hence it is hypothesized that this approach promotes the development of CT (Thomas, 2011). In an effort to validate the application of the GR test[2]—and thus the effectiveness of the course redesign—a statistical analysis helped investigate the relative merit of the test.

   We asked students who were registered in the course during the 2012 and 2013 (n= 55) iterations to voluntarily participate in the WGCTA assessment. The two cohorts were combined (n=34; 3 were later excluded for failing to complete all course requirements) and compared to students from the 2011 iteration of the course (n=22) to test whether there were any differences between the two populations.

   The test was administered during a regularly scheduled tutorial session, which had been assigned as an open date (neither the course instructor nor the course TA were in attendance). A member of the academic skills centre who was not directly affiliated with the course delivered the test. Student performance on this test was not shared with the course instructor. At the end of the course and after all grades had been submitted, the data was analyzed using SPSS and subjected to a regression analysis designed to determine if there was any correlation between the test scores on the GR and WGCTA tests. Data from 2012 and 2013 were pooled and students' final scores on the GR and assignments were compared with the WGCTA performance scores. In order to verify the capacity of the GR test to serve as a measure of CT performance, the WGCTA was examined for correlation with the GR test using a paired samples T-test. A probability plot was constructed to test normality of the scores for each variable. Given the small "n" for students

who participated in the WGCTA in each of the years (11 in 2012, 23 in 2013) it was necessary to determine if data from the test groups could be pooled. Application of Levene's test to examine whether there was any statistically significant variance between the two test group populations was undertaken and followed by an independent sample T-test. (See results section for p-values and other relevant data.)

In addition to examining the student performance of CT, we also evaluated the impact of the course changes on perceptions of the course instructor through the use of independently[3] administered SOS using a Likert scale questionnaire scored out of a total of seven. For the specific question about the instructor ("The instructor created a course atmosphere that was conducive to my learning."), an average score ranges from 5.6 to 5.8 with a strong score reading from 5.9 to 6.0. A score above 6.0 is regarded as exceptional; scores under 5 are considered weak. There is also an opportunity for students to offer personal commentary.

RESULTS

Table 1 presents the statistical significance of the course assessment tools as they relate to their capacity to measure CT by examining the relative performance of a group of students (31) enrolled in the course whose 2012 and 2013 data was pooled. Note that the grouping of the 2012 and 2013 WGCTA data is possible as there was no variance in between the two iterations of the WGCTA test populations (p-value of 0.158 & 0.477). In addition a paired samples T-test examining the correlation between the WGCTA and GR tests has a p-value of < 0.001 with a moderate positive correlation of 0.388, suggesting that the GR test is a viable alternative measure of CT in the absence of the WGCTA.

There also appears to be a trend in student performance according to test scores. For example, students in the test group exhibit a decline in their MC test scores from 65% in the control group down to 56% for the test group (p-value 0.048) while simultaneously displaying an overall increase in the total score for the final exam, moving from 65% to 69% (p-value 0.118) respectively when GR and MC scores are combined. Although this upward trend is not statistically significant, it might suggest an attempt on the part of students to dedicate more time in their preparation for GR test questions, which could be interpreted as a commitment to explore deeper understanding. At the very least, the

Table 1. Justification for the pooling of 2012 and 2013 WGCTA student performance data as given by the absence of any statistically significant difference in student performance (p-values of 0.158 & 0.477). Listing of mean scores (as %) and standard deviations (SD) along with confidence intervals and relevant p-values for assignments and summative assessment tasks including exams using standard MC questions and those that combined MC and GR questions.

| ATTRIBUTE | CONTROL GROUP | | TEST GROUP | | 95% CONFIDENCE | |
	MEAN	±SD	MEAN	±SD	INTERVALS	P-VALUE
Assignment # 1	78	±7.4	74	±13.7	−1.7–10.1	0.158*
Assignment # 2	72	±9.1	69	±14.6	−4.3–9.1	0.477*
Exam MC vs. MC	65	±10.6	56	±17.2	0.9–16.7	0.048
Exam MC vs. GR+MC	65	±10.6	69	±9.0	−9.9–1.1	0.118

* (asterisk) *refers to values reflecting no variance between means as determined by the independent T-test for equality of means.*

Table 2. Student performance data compared across the six latest iterations of the course.

ITERATION	ENROLLED	COMPLETED	% COMP.	D & FAILS	% OF D & FAILS	A, B VS. C, D, & F RATIO	#SOS SCORE	SOS RESP.	% RESP.
A (2010) ★	180	153	85%	24	15.69	85/153 = 56%	4.8	70	45.45
B (2010) ★	98★	68★	69%★	9★	13.24★	40/68 = 59%★	5.7★	44★	64.71★
C (2011)*	150*	103*	69%*	27*	26.21*	49/103* = 48%	NA*	NA*	NA*
D (2011) ★	65★	22★	34%★	3★	13.63★	15/22 = 68%★	6.5★	22★	100★
E (2012)	37	30	81%	1	3.33	15/30 = 50%	6.1	20	66.67
F (2013)Δ	33	25	76%	2	8.00	12/25 = 48%	NA	NA	NA

* (asterisk) represents session not taught by the instructor undertaking the investigation.

★ (star) denotes summer session. All other sessions are winter. Italics as in row C denotes an instance of the course being taught by another instructor and therefore a return to pre-2010 format (i.e, exam only).

BOLD denotes courses where some intervention was applied. Note that in the Summer 2010 ★ iteration (B), a revamped but incompletely revised syllabus was distributed, but only one academic writing tutorial was included, and there was no focus group intervention or other student support, other than traditional office hours.

\# indicates SOS scores are out of a maximum of 7.

Δ means SOS scores are not available for this iteration. Surveys were solicited online for the first time, resulting in a very limited response rate (i.e., 8/30 responded). Note that the current literature (Adams & Umbach, 2011) seems to suggest that a decline in participation is normal at the outset for an on-line survey.

Table 3. Comparison of course structure, pre-intervention vs. post intervention models

PRE-INTERVENTION (2010)	POST INTERVENTION (2011-13)
Syllabus informative but insufficiently descriptive.	Syllabus fully aligned, linking course objectives to assessment tasks and course materials.
Standard multiple choice testing combined with short answer and essay questions.	Graded response (GR) testing combined with short answer and essay questions.
Assignment essays include outline and general rubric. Multiple drafts not available.	Assignment essays include outline, detailed rubric, and use of multiple drafts aimed at encouraging student self-reflection.
Tutorials aimed at content specific support.	Tutorials redesigned to encourage student feedback through modeling of CT behaviours.
Student opinion not sought out formally, other than through the university's SoS.	Student opinion sought formally through embedded focus group with an emphasis on reflection, as well as SoS.
Instruction tending to focus on "what" or content acquisition.	Instruction focusing on "why" or justification of rationale.

findings suggest that students may be spending somewhat less time memorizing content-specific statements and committing more time to exploring concepts that require greater reflection. An additional finding as provided by the SOS scores was the fundamental shift in the response of the students from generally negative to significantly positive. Table 2 above compares the overall rating of the instructor from the pre-intervention in 2010 to the 2012 iteration. Included in the table are the respective SOS scores and the percent of respondent rate for these surveys.

Additionally, prior to the redesign, students had been quite critical of the course expectations, offering frequent statements such as "too much information . . . delivered at a stifling pace." By 2012, the comments had changed dramatically as evidenced by this comment: "Professor is very enthusiastic and simplifies a complicated concept. Focuses on the key concepts for testing information. Expects students to understand and not just memorize important facts. I really enjoyed his class." Given the aim of the course re-design, a comment of this type represents an opportunity to validate, albeit in a qualitative manner, student willingness to recognize their responsibility in their own learning. In our opinion, this view suggests the student's acceptance that it is insufficient to collect facts and that there is greater value in determining how or when to apply these facts.

CONCLUSION

Collaborating with the educational developer on a course design through the lens of SoTL principles of good practice allowed us to shift toward a learning centered environment that provided the students with opportunities to develop their CT skills. With assessment tools that can effectively validate the course's CT goals through a standardized CT test like the WGCTA or a discipline-specific tool like the GR test, we can not only support the development of CT but also verify its presence. Effectively, we have developed a step-wise process (Table 3) that can ensure appropriate alignment between course outcomes and course delivery.

We acknowledge that there may be faculty resistance to implementing the necessary changes, but the benefits significantly outweigh the costs (Gardiner, 1998). Where resources are available, it is advantageous for university faculties to place a high priority on the development of pedagogically sound practices that will benefit both the student and the instructor. Much of the SoTL literature (Bernstein 2008; Felten, 2013; Trigwell, 2013) emphasizes the need to share in teaching and learning projects (specifically SoTL, by also by extrapolation, course redesign), so that there is a greater sense of objectivity and pedagogical value when evaluating the merits of any given instructional strategy. A collaborative effort is, therefore, ideal to use in course redesign processes, especially those that apply the principles of good practice in SoTL.

Michael deBraga is a lecturer supporting mathematics and science in the Robert Gillespie Academic Skills Center and teaches in the Biology Department at University of Toronto Mississauga (CANADA).

Cleo Boyd is a senior lecturer and educational developer in the Robert Gillespie Academic Skills Center at University of Toronto Mississauga (CANADA).

Shahad Abdulnour is an academic strategist in the Robert Gillespie Academic Skills Center and teaches in the departments of Historical Studies and Biology at University of Toronto Mississauga (CANADA).

NOTES

1. Here is an example from the WGCTA Short Form test booklet (Watson & Glaser, 2006, pg.3). "Assign to each statement a designation as to whether it is TRUE, PARTLY TRUE, INDETERMINATE, PARTLY FALSE, or FALSE: Two hundred students in their early teens voluntarily attended a recent weekend student conference in a Midwestern city. At this conference, the topic of race relations and means of achieving lasting world peace were discussed, because these were the problems the students selected as being most vital in today's world.

    1. As a group, the students who attended this conference showed a keener interest in the broad social problems than do most other students in their early teens.
    2. The majority of the students had not previously discussed the conference topics in their schools.
    3. The students came from all sections of the country.
    4. The students discussed mainly labor relations problems.
    5. Some teenage students felt it worthwhile to discuss problems of race relations and ways of achieving world peace."

2. The WGCTA has been shown to effectively measure CT skills. A regression analysis on a sample of self-selected students was undertaken to evaluate whether there was a correlation between the performance of the students on the GR test and course assignments and the WGCTA (see results section).

3. Student opinion surveys (SOS) are administered through the university for every course.

REFERENCES

Abualrub, I., Karseth, B., & Stensaker, B. (2013). The various understandings of learning environment in higher education and its quality implications. *Quality in Higher Education 19(1),* 90-110. doi: 10.1080/13538322.2013.772464

Aldridge, J.M. & Fraser, B.J. (2000). A cross-cultural study of classroom learning environments in Australia and Taiwan. *Learning Environments Research 3,* 101-134.

Aldridge, J.M., Dorman, J.P., & Fraser, B.J. (2004). Use of multitrait-multimethod modelling to validate actual and preferred forms of the technology-rich outcomes-focused learning environment inventory (Troflei). *Australian Journal of Educational & Developmental Psychology. 4*, 110-125. http://www.newcastle.edu.au/journal/ajedp/

Beachford, M.R. & Beachford, J.C. (2010). Critical-thinking pedagogy and student perceptions of university contributions to their academic development: *International Journal of an Emerging Transdiscipline, 13,* 53-71. http://www.inform.nu/Articles/Vol13/ISJv13p053 -071Beachboard548.pdf

Bernstein, D.J. (2008). Peer review and evaluation of the intellectual work of teaching. *Change: The Magazine of Higher Learning, 40(2),* 48-51. doi: 10.3200/CHNG.40.2.48-51

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education 32,* 347-364.

Biggs, J. (1999). What the student does: teaching for enhanced learning. *Higher Education Research & Development 18(1),* 57-75.

Biggs, J. & Tang, C. (2007). *Teaching for quality learning at university: what the student does* (3rd ed.). New York NY: McGraw-Hill Education.

Blair, E. (2013). The challenge of contextualising the scholarship of teaching and learning. *Teaching & Learning Inquiry, 1(1),* 127-130. doi: 10.1353/iss.2013.0000

Brookfield, S.D. (1997). Assessing critical thinking. *New Directions for Adult and Continuing Education, 75,* 17-29.

Covington, M. (1984). The motive for self-worth. *Research on Motivation in Education, 1*, 77-113, Orlando, Florida, Academic Press.

Entwistle, N. (2000). Promoting deep learning through teaching and assessment: conceptual frameworks and educational contexts. *Paper presented at TLRP [Teaching & Learning Research Programme], Leicester, November 2000,* 12pgs.

Felten, P. (2013). Principles of good practice in SoTL. *Teaching & Learning Inquiry, 1(1),* 121-125. doi: 10.1353/iss.2013.0013

Froyd, J. & Simpson, N. (2010). Student-centered learning addressing faculty questions about student-centered learning. Texas A&M University. http://ccliconference.org/files/2010/03 /Froyd_Stu-CenteredLearning.pdf

Gardiner, L.F. (1998). Why we must change: the research evidence. *The National Education Association Higher Education Journal (Spring 1998),* 71-88.

Ghaye, T. (2011). *Teaching and learning through reflective practice: a practical guide for positive action* (2nd ed.). New York, NY: Routledge.

Haas, P.F. & Kelley, S.M. (1998). Coping with faculty resistance to teaching critical thinking. *College Teaching, 46(2),* 63-67.

Hounsell, D., McCune, V., Litjens, J., & Hounsell, J. (2005). Enhancing teaching-learning environments in undergraduate courses: subject overview – Biosciences, December 2005. The ETL project: Universities of Edinburgh, Coventry and Durham. 60pgs. http://www .etl.tla.ed.ac.uk//docs/BiosciencesSR.pdf

Keeley, S.M., Shemberg, K.M., Cowell, B.S., & Zinnbauer, B.J. (1995). Coping with student resistance to critical thinking. *College Teaching, 43(4),* 140-145.

Kyndt, E., Dochy, F., Struyven, K., & Cascallar, E. (2011). The perception of workload and task complexity and its influence on students' approaches to learning: a study in higher education. *European Journal of Psychology Education, 26*, 393-415. doi: 10.1007/s10212-010-0053-2

Leshowitz, B., DiCerbo, K.E., & Symington, S. (1999). Effective thinking: an active-learning course in critical thinking. *Current Issues in Education, 2(5)*, 1-12.

Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: implications for theory and practice. *Studies in Higher Education 27(1)*, 27-52. doi:10.1080/03075070120099359

Magnussen, L., Ishida, D., & Itano, J. (2000). The impact of the use of inquiry-based learning as a teaching methodology on the development of critical thinking. *Journal of Nursing Education 39(8),* 360-364.

Mathews, S.R. & Lowe, K. (2011). Classroom environments that foster a disposition for critical thinking. *Learning Environments Research 14(1),* 59-73. doi: 10.1007/s10984-011-9082-2

Morrison, H. (1997). Information literacy skills: an exploratory focus group study of student perceptions. *Research Strategies, 15(1),* 4-17.

Prideaux, D. (2003). ABC of learning and teaching in medicine: curriculum design. *British Medical Journal 326(7883),* 268-270. doi: http://dx.doi.org.myaccess.library.utoronto.ca/10.1136/bmj.326.7383.268

Seifert, T. (2004). Understanding student motivation. *Educational Research, 46(2)*, 137-149. doi: 10.1080/0013188042000222421

Scott, W., Lisagor, T., & Marachi, R. (2009). Learning centered universities: the changing face of higher education. *Journal of Faculty Development (23)1,* 14-23.

Taylor, M.L. (2006). Generation NeXt comes to college: 2006 updates and emerging Issues. *A Collection of Papers on Self-Study and Institutional Improvement - Arkansas State University-Beebe, 2,* 48-55. https://www.msu.edu/~jdowell/pdf/Millennials.pdf

Thomas, T. (2011). Developing first year students' critical thinking skills. *Asian Social Science 7(4),* 26-35. doi: 10.5539/ass.v7n4p26

Trigwell, K. (2013). Evidence of the impact of scholarship of teaching. *Teaching & Learning Inquiry, 1(1),* 95-105. doi: 10.1353/iss.2013.0004

Tsui, L. (2002). Fostering critical thinking through effective pedagogy: evidence from four institutional case studies. *The Journal of Higher Education 73(6),* 740-763. doi: 10.1353/jhe.2002.0056

Watson, G. & Glaser, E.M. (2006). Watson-Glaser critical thinking appraisal (short form). Pearson Education Inc. San-Antonio, USA.

APPENDIX 1

Lecture and tutorial content

1) How prepared were you prior to the start of this course? For example, how confident were you about meeting the course expectations?

2) Now that you have passed the halfway point of the course, with respect to your performance, how confident are you that you will successfully meet the course expectations?

3) Do you feel that the course pre-requisites have sufficiently prepared you for this course?

4) Are course tutorials helpful in preparing you for your assignments? What do you find most helpful? What would you like to see added or deleted?

5) What did you find most challenging about your most recent assignment?

6) Regarding the visit from the science librarian, have you experienced librarian visits before?

7) How did this visit differ from others you have had?

8) Was the visit by the librarian helpful in presenting you with an overview of the available resources?

Instructor assessment

9) Does the course syllabus adequately lay out the course requirements and the necessary resources? For example, is the course content on Blackboard laid out in an easy to access manner (i.e., course information folder, course lecture folder, etc.)?

10) Do you find the narrative style of your instructor enjoyable? Why or why not?

11) Did the ROM (Royal Ontario Museum) visit provide you with an opportunity to interact with your peers and with your instructor in a less formal manner? Was this opportunity beneficial?

12) Do you appreciate the interactive nature of how the course material is delivered? In other words, do you find that concepts are presented without necessarily having a resolution?