Higher Education Quality Council of Ontario

An agency of the Government of Ontario

Building Capacity to Measure Essential Employability Skills:
A Focus on Critical Thinking

Gary Kapelus, Nancy Miyagi, Valerie Scovill
George Brown College

Cite this publication in the following format:

Kapelus, G., Miyagi, N., Scovill, V. (2017) *Building Capacity to Measure Essential Employability Skills: A Focus on Critical Thinking.* Toronto: Higher Education Quality Council of Ontario.

Acknowledgements

Executive Summary

This report is the culmination of a three-year research project conducted by George Brown College (GBC). As a member of the Learning Outcomes Assessment Consortium, sponsored and funded by the Higher Education Quality Council of Ontario (HEQCO), this project responds to HEQCO's request for colleges and universities to develop, implement and share new assessment tools that "measure and validate the attainment of these generic learning and cognitive skills."[1]

In this project, we focused on critical thinking (CT), with the goal of addressing a fundamental question: How do we measure student learning of this essential employability skill during the course of a program of study?

Specifically, we wanted to know:

1. What are the measurable elements of critical thinking relevant to postsecondary education?
2. Do our existing assessment tools allow us to measure critical thinking?
3. Are students currently learning and demonstrating critical thinking skills in their coursework?
4. Are students aware of their own critical thinking skills?
5. What would be the elements of a generic assessment tool for critical thinking that could be adapted for use in any course in the college?

To explore these questions, the project team developed a three-phase methodology that directly involved faculty and students from a broad cross-section of the college, with the following objectives:

*   To develop and pilot test practical and relevant new curriculum resources and a validated assessment tool for the learning, practice and demonstration of the essential employability skill (EES) of critical thinking within a course.
*   To design these resources to enable adaptation and application by faculty within any program curriculum across the college and, ultimately, within any postsecondary curriculum.
*   To pilot a collaborative developmental process for one EES that can be applied to the development of similar tools and resources for other EES.

Key Findings and Conclusion

Challenges in Developing a Generic Assessment Tool

We learned that critical thinking was not explicitly being addressed or assessed within existing college English courses and rubrics. We learned that assessing critical thinking was complicated or obscured by the quality of students' English language skills and that faculty required additional training to separate assessment of critical thinking and language. We learned that faculty had many different interpretations of critical thinking and how it could be demonstrated in student assignments within their unique contexts and

---

1 HEQCO-RFP-006: HEQCO Learning Outcomes Assessment Consortium (Universities and Colleges); December 3, 2012

that considerable discussion was required to arrive at a consensus about the specific constructs to be included in a generic tool and how to best describe generic performance levels for each. Our blind marking rounds demonstrated the challenges of finding appropriate and relevant samples of student work to test the evolving rubric when available student samples were based on previous assignments that did not explicitly incorporate the constructs being assessed by the rubric.

Evaluating Only What Has Been Taught and/or Practised

Faculty feedback from the project volunteers and others confirmed that developing an assessment tool on its own provided no benefit unless all the elements to be assessed were also clearly incorporated into the curriculum. Through this project, we have identified a collaborative, inclusive process and the pedagogical resources needed to meet that objective.

This project has resulted in the following products:

A CT Assessment Rubric

We identified a core set of CT constructs, rooted in the literature, that are viewed as most relevant by college faculty. Based on these, we developed a reliable CT assessment rubric that provides the college with an excellent starting point from which to assess students' CT skills consistently across the curriculum. Improving the reliability of the rubric is a work in progress.

Critical Thinking: Learning, Teaching and Assessment — A Teacher's Handbook

With the assistance of a curriculum specialist, a faculty working group incorporated the core set of CT constructs into their existing course curriculum, impacting on their course outlines, what was discussed and taught, and how this was assessed in assignments. Along with the CT rubric, practical suggestions and real examples for incorporating the CT constructs and how to use the rubric, taken directly from the work of this faculty group, were fashioned into a valuable handbook for teachers, and distributed widely. Faculty were encouraged to use this resource in their own classroom.

Essential Employability Skills: A Teacher's Guide

We created an additional teacher resource with practical examples and suggestions for incorporating all 11 EES into any college curriculum. This guidebook does not, however, contain rubrics for each EES.

A Pilot-tested, Collaborative Process for Developing Measurement Tools and Associated Curriculum Resources that Can Be Utilized for Other EES

The project was designed to enable a collaborative, inclusive developmental process to unfold, engage a representative group of faculty volunteers and foster significant change in curriculum and teaching practices and result in the products described above.

Moving forward, George Brown College will continue to build on the existing work and consider the lessons learned and recommendations provided by the project team.

Table of Contents

Quotes

"I used it to actually help clarify assignment instructions. Before this project I did not have a concrete way of deciding when students were being critical. I had an idea of what I was looking for but found it difficult to communicate this to the students." — General Education faculty participant

"As a result of this project and in consultations with [name]*, I have slightly modified my course outcomes to better include a CT focus. I am certainly more aware of critical thinking skills myself, and I am still fine-tuning my delivery and inclusion of CT skills within my online courses."* — General Education faculty participant

"The greatest challenge is not the constructs or the rubric in its entirety, but my own level of complacency when attempting to judge if critical thinking was happening. I guess what I'm saying is that I was hesitant to implement the rubric and I felt like we were trying to put a square peg in a round hole. But once accepting that the tool had to evolve with my teaching and assessments, I began to see its worth and value." — General Education faculty participant

"As a result of this blind marking assignment, I am focusing so much more on the critical thinking skills you have raised in this rubric and, therefore, I will have a better idea of how/whether the rubric reflects the teachable skills at the end of this term!" — Blind marking faculty volunteer

"I found that I started to think about how I was teaching differing points of view and the type of information I was including as examples. I think I really became more aware of my own biases and perspective in my teaching." — General Education faculty participant

"[The rubric] has left me wondering how I might change my teaching to address some of these skills (specifically, critiquing contradictory evidence)." — Blind marking faculty volunteer

1. Introduction

This report is the culmination of a three-year research project conducted by George Brown College (GBC). As a member of the Learning Outcomes Assessment Consortium, sponsored and funded by the Higher Education Quality Council of Ontario (HEQCO), this project responds to HEQCO's request for colleges and universities to develop, implement and share new assessment tools that "measure and validate the attainment of these generic learning and cognitive skills."[2]

In Ontario community college curricula, these generic learning and cognitive skills are expressed as essential employability skills (EES) and their associated learning outcomes. Six categories of EES have been defined by the Ontario Ministry of Advanced Education and Skills Development[3] (see Appendix A). These EES, along with vocation-specific learning outcomes and educational breadth requirements, help to define community college program curricula.

The general challenge for Ontario colleges has been to translate the broadly worded EES and their 11 associated learning outcomes into more clearly defined, context-specific and measurable learning outcomes that are relevant to a particular course.

To address this challenge, the intent of this project is to develop and pilot practical new curriculum resources and assessment tools for the learning, practice and demonstration of EES within a course. The objective is to provide tools, which can be adapted and applied by faculty within any program curriculum across the college and, ultimately, within any postsecondary curriculum.

In this project, we have focused on one particular EES category, namely critical thinking (CT), with the goal of addressing a fundamental question: How do we measure student learning of this essential skill during the course of a program of study?

The specific focus on CT skills responds to insights obtained from our graduates and aligns with the job-ready cognitive skill requirements previously identified by employers and our industry partners.

Specifically, we want to know:

1.   What are the measurable elements of critical thinking relevant to a postsecondary education?

2.   Do our existing assessment tools allow us to measure critical thinking?

3.   Are students currently learning and demonstrating critical thinking skills in their coursework?

---

2 HEQCO-RFP-006: HEQCO Learning Outcomes Assessment Consortium (Universities and Colleges); December 3, 2012
3 Ministry of Advanced Education and Skills Development (formerly Training, Colleges and Universities), "Appendix B: Essential Employability Skills," Framework for Programs of Instruction. Issued April 1, 2003; revised, April 1, 2005

4.   Are students aware of their own critical thinking skills?

5.   What would be the elements of a generic assessment tool for critical thinking that could be adapted for use in any course in the college?

To explore these questions, the project team developed a three-phase methodology that directly involves faculty and students from a broad cross section of the college:

- Phase 1: (i) literature review and (ii) identification of current practices within college English courses (2013–14).

- Phase 2: development and pilot testing intervention EES curriculum and assessment tools for critical thinking within the General Education/Liberal Arts environment (2014–15).

- Phase 3: knowledge transfer (2015–16).

2.   Literature Review and Identification of Current Practices Within College English Courses

A literature review was conducted at the beginning of the project (2013–14) to gain insight into a range of topics relevant to this research study. Specifically, the review focused on addressing three research questions:

- What are the common measurable elements or constructs of critical thinking?

- How is critical thinking generally measured? What should be considered when developing rubrics specific for critical thinking?

- Can critical thinking skills be self-assessed?

Findings from the review are incorporated throughout this report. For additional details, refer to Appendix B.

In Phase 1, researchers referred to the literature to identify current practices related to the teaching and assessment of CT, including commonly accepted elements or constructs that define CT skills. Consultations with college English (COMM) faculty on their current practices further supplemented the review. COMM faculty were specifically engaged at this stage both because of their unique focus on the development of students' expressive language skills and the breadth of their teaching assignments across, and therefore reflection of, all programs at the college. This work informed revisions to our research approach in Phases 2 and 3.

During semi-structured faculty focus group consultations in Phase 1, it was revealed that professors do not follow a standard definition of CT and therefore, the elements of this EES are loosely defined. However, this finding is not limited to the GBC context. The literature review confirmed that there is no single, widely accepted definition of CT. Definitions range from in-depth descriptions of declarative knowledge, procedural knowledge and metacognition (Kurfiss, 1988) to the habits undertaken when thinking critically (Marzano et al, 1993) and a simple broad description that critical thinking is "the kind of thinking that professionals in the discipline use when doing the work of the discipline" (Taylor, 2004). Though definitions differ, most definitions share some commonalities. Many agree that critical thinking is a skill or process and involves attitudes, habits, values and behaviour (Rickles et al., 2013). As a result of these variations, there are different constructs of CT.

In order to reach some level of consensus at this early stage of the project, COMM faculty were provided several definitions of critical thinking (Brookfield, 1997; Bok, 2006; Facione, 1990; American Association of College and Universities, n.d.). In their focus groups, faculty were first asked how they currently taught and assessed CT skills in their courses, and then asked to reflect on these definitions and their own understanding of CT, and to provide recommendations on the specific elements of this skill that we should measure. COMM faculty favoured two definitions that were thought to be most relevant and most closely reflective of both their teaching practice and their expectations for students. The first states that critical thinking is "purposeful, self-regulatory judgment, which results in interpretation, analysis, evaluation, and inference as well as explanation of the evidential, conceptual and methodological considerations on which a judgment is based" (Facione, 1990). The second, simpler definition came from the Association of American College and Universities (AACU). The AACU CT rubric states that "critical thinking is a habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion."

Based on current practice, COMM faculty considered at minimum, the following elements or constructs of critical thinking to be important in relation to typical college-level assignments:

- Expressing critical thinking through written assignments that respond to articles that mostly take a stand or represent a position.
- Identifying a problem and finding a solution to a problem to encourage students to formulate different solutions to problems or find different perspectives on an issue, rather than thinking there are only two sides.
- Evaluating and properly citing sources because the COMM1007 courses may be the only place — or one of the few places — in a program where students will explicitly learn this, but they will be expected to apply it in other courses.
- Supporting a position taken to teach students the value of effective research, a skill which will also be utilized in the rest of their programs. Included in this category are the proficiencies of recognizing one's own biases and acknowledging other viewpoints and contradictory evidence in order to provide balanced research.

- Because the COMM faculty value the effectiveness of strategically chosen language for both the recognition of a position taken in a piece of writing and the expression of CT skills when responding to a position, language proficiency was thought to be important.

Upon reviewing the existing college English rubric tool against the above constructs, the COMM faculty working group concluded that the existing rubric did not adequately reflect these constructs and would not be useful in measuring these within typical English course assignments. Furthermore, the working group identified the challenges in marking both for communication (using existing marking rubrics for college English courses) and for critical thinking. Ultimately, the project team determined that the original objective to focus on two related essential employability skills, both communication and critical thinking, was too broad and complex for this study. As a result, we narrowed the focus of our work in Phases 2 and 3 to making critical thinking more explicit within the curriculum. It was agreed that a provisional rubric tool should be developed for this purpose with input from the COMM faculty. Although there are limitations with having these constructs defined through the lens of COMM faculty only, it was viewed as a necessary first step in conceptualizing a set of elements that, with additional faculty input in Phases 2 and 3, could lead to a standard college-wide definition of CT.

3. Development of a CT Assessment Tool and Associated Curriculum Interventions

The work in Phases 2 and 3 entailed two major components: (i) development, validation and pilot testing of a common CT assessment tool (rubric); and (ii) development and then incorporation of new teaching and learning strategies into course curricula that specifically address CT skills and utilize the standard rubric.

This section summarizes the work completed between February 2014 and December 2015, including research methods, results and analyses. Additional analysis and survey forms can be found in the appendix section.

3.1 CT Rubric

Development of CT Rubric

Several tasks were completed to develop and validate a new CT rubric, starting with the development of a provisional rubric first, followed by three rounds of blind marking and faculty debriefings. Results from the blind marking rounds, which also sought feedback from General Education/Liberal Arts faculty about their experience with using the rubric, and post-marking debriefings informed revisions to the rubric. Refer to Appendix C for a copy of the final rubric, George Brown College CT Rubric, and Appendix D for the Critical Thinking — Blind Marking Recording Form.

The following describes the activities the project team undertook to develop and validate the CT rubric. For results, refer to Table 1.

CT Rubric Version 1

Following the literature review (Appendix B) and consultations with COMM faculty in Phase 1 of the project, the project team proceeded to develop a provisional CT assessment rubric designed specifically for marking assignments within a college English course. The initial version of the assessment rubric identified five CT constructs (labeled as criteria) and included an additional communication checklist to allow markers to indicate an overall sense of the vocabulary and language being used to communicate critical thinking in written assignments. To assist the marker in identifying the level to which a student demonstrates achievement of each of the CT elements, the assessment tool provided three rating levels: (3) excels; (2) effectively meets the criterion; (1) falls short/needs improvement. The additional communication checklist used a slightly different rating scale, reflecting English teachers' expectations for vocabulary and language structures used in the paper.

Blind marking #1:

Eleven college English professors volunteered to assess each sample paper using the provisional CT rubric, between February 18 and March 6, 2014. Samples of recently graded college English written assignments were solicited from the college English faculty that implicitly required students to use CT skills, even though these assignments were neither specifically intended nor graded for critical thinking. Five representative anonymous student papers from two different types of assignments were selected. Markers were provided with the provisional rubric and sample assignments to be marked. Marking was done in a classroom setting, with markers assessing the papers on their own. The sessions lasted about two hours and were led by the curriculum specialist from the project team. It was assumed that, as the rubric tool contained detailed scale descriptors and corresponding rating levels (1 to 3), it was self-explanatory. Hence, no detailed explanation or instructions were provided to volunteers about using the rubric.

CT Rubric Version 2

Based on results from the first blind marking and in consultation with the COMM faculty working group from Phase 1 and General Education/Liberal Arts faculty participants recruited for the project in Phase 2, the project team developed version 2 of the CT rubric. The second rubric included six CT constructs and four levels of mastery or rating scale: (4) exemplary; (3) meets expectations; (2) needs improvement; (1) missing information.

Blind marking #2:

The second blind marking round was conducted in December 2014 using six randomly selected anonymous student samples obtained from three General Education/Liberal Arts courses that had not been exposed to curriculum changes related to CT, the control variable. For this marking round, sample written assignments were sought that specifically purported to demonstrate CT skills, although not necessarily inclusive or reflective of the six specific CT constructs in the CT rubric. Of the six assignments ultimately chosen for blind marking in this round, there was some inconsistency between the original assignment instructions (i.e. what the students were asked to demonstrate) and the CT rubric (what the blind markers were asked to assess). For example, some assignments, which purported to require critical thinking, did not reflect this in the written instructions to students. Some assignments did not reflect all of the constructs in the CT rubric.

Despite these drawbacks, a decision was made to instruct blind markers to assess evidence of students' CT skills against the six CT constructs. Eighteen faculty (six General Education/Liberal Arts faculty participants and 12 faculty volunteers from other areas of the college) were asked to mark the six student samples. Markers were provided with the CT assessment rubric, student samples, original assignment instructions (what students had actually read) and a blind marking recording form with instructions for faculty to record their answers.

CT Rubric Version 3

Based on results from the second blind marking round and faculty feedback, the subsequent revision of the rubric included six CT constructs that were more specifically descriptive than those of the previous rubrics and four levels of mastery/rating scale (without a numeric value): Exemplary/exceeds expectations; meets expectations; below expectations; inadequate.

Blind marking #3:

The third and final blind marking round was conducted in April and May of 2015. In this round, anonymous student samples were randomly selected from three General Education/Liberal Arts courses where curriculum changes had been implemented and assignments had been developed in alignment with the revised CT rubric. The curriculum specialist from the project team also compared assignment instructions (what students were asked to do) against the rubric (what blind markers were asked to assess) and confirmed that one of the assignments did not meet one criterion. As a result, markers were asked to omit it from their assessment. Twenty-two faculty — six General Education faculty participants and 16 faculty volunteers from other academic departments — were invited to participate in blind marking. Of this group, 21 professors fully marked all samples and recorded their answers in the blind marking recording form.

Blind Marking Results

Results from the blind markings were analyzed using two measures: inter-rater agreement and inter-rater reliability. Given that one of the general goals of this project is to develop the initial assessment tool (the CT

rubric) for measuring skill acquisition in critical thinking across the college, it was felt that assessing both inter-rater agreement (IRA) and inter-rater reliability (IRR) of blind markings would provide useful information to inform future changes to the assessment tool. While agreement and reliability are used interchangeably in research studies, each represents a different concept. "Agreement is defined as the degree to which scores/ratings are identical, whereas reliability relates to the extent of variability and error inherent in a measurement" (Gisev, Pharm, Bell, and Chen, 2013, p. 331). IRA was analyzed using percentage of exact agreement within each level of mastery for each criterion. We report the highest proportion of agreement reached among markers. IRR was analyzed using intra-class coefficient (ICC), mixed model for consistency. The analysis was done using SPSS 20.0. As a rule-of-thumb, we used 75% as the minimum acceptable score for absolute agreement (IRA) and .8 for IRR (Graham, Milanowski and Miller, 2012).

This section provides results from the three blind markings as well as a summary of the feedback received from General Education and Liberal Arts faculty. Table 1, at the end of this section, summarizes the IRA and IRR scores.

Blind marking #1

Of the 11 faculty volunteers, six fully marked all five student samples using the provisional rubric. However, a number of markers either skipped the communication checklist component or made multiple selections of achievement level rather than choosing one as intended. As a result, the confusing and missing data from the communication checklist item were removed from the data analysis. Therefore, the final data analysis involved six markers, five samples and five CT criteria. As shown in Table 1, results indicate a low level of IRA overall, with absolute agreement per criterion ranging from 50% for Criteria 1 (Identification of issue/problem) to 73% for Criteria 5 (Conclusions), below the 75% benchmark. The ICC for individual criterion ranged from .51 (Criteria 2: Support for position taken) to .82 (Criteria 3: Language used in response). Although results reflect the scores provided by a very small subset of COMM faculty volunteers, it became evident that the provisional CT rubric needed significant revisions.

Following the blind marking session and after the scores were tabulated, the curriculum specialist debriefed the COMM faculty group. The discussion focused on concerns about the wording, clarity and applicability of the CT criteria used in the tool, challenges in applying these in actual marking, reasons for variation among blind markers in applying the CT assessment tool, and the appropriateness of the communication checklist as part of the assessment tool. These discussions revealed a number of issues:

- **Unclear criteria descriptors**: The initial wording of the descriptors for the five CT assessment criteria was not as self-evident or clear as originally thought, and allowed for considerable misinterpretation by markers when they attempted to apply these in marking papers.

- **More than one CT element in the same construct:** Markers also identified that two of the CT criteria were difficult to apply because they were actually composites of more than one construct that needed to be separated out.

- **Communication and language:** The difficulties identified by markers using the communication

checklist and the potential overlap with the Language Used criterion, resulted in dropping the communication checklist from the tool. It was argued that the existing Language Used criterion should be retained, again reflecting the inherent desire of English faculty to assess the quality of language in expressing critical thinking.

- **Lack of instructions on how to use the rubric:** Lack of instructions to faculty on the use of the rubric may have contributed to incomplete markings for every criterion in all five sample assignments, which resulted in the final selection of only six faculty markings for analysis purposes.

- **Separating CT from written communication skills:** The discussion among the faculty group also identified the inherent challenges for a marker in mentally separating the demonstration of critical thinking in a written paper from the way in which these ideas are communicated; that is the quality of English language used. For a postsecondary environment in which there are many students whose first language is not English and some native English speakers with weak English language written skills, a student's English capabilities (for example, choice, clarity and sophistication of language, vocabulary and grammar, organization of ideas, etc.) may complicate and in some cases obscure their demonstration of CT skills in a written assignment. It may also be that sophisticated English language written skills might mask less well developed CT abilities in a written assignment.

In conclusion, all of the above may have contributed to variability in the blind marking scores and consequently, a low inter-rater agreement and intra-class correlation co-efficiency scores, overall, which make the initial assessment tool an unreliable instrument for assessing CT skills.

On the basis of these discussions, the CT assessment tool was revised. The next phase of development of the CT assessment tool utilized a broader group of faculty volunteers and perspectives, with a specific focus on General Education and Liberal Studies representing Arts and Humanities, Science and Technology, and Social Sciences.

Blind marking #2

Of the 18 faculty volunteers, 13 professors fully marked all six student samples. Results from the second blind marking round indicate that IRA scores were below the target of 75%, with the highest absolute agreement per criterion ranging from 45% for Criteria 3 (Analysis of support) to 55% for Criteria 6 (Conclusions). The ICC for individual criterion ranged from .83 (Criteria 2: Choice of support sources) to .97 (Criteria 6: Conclusions) well above the target of .80. Compared to the first blind marking, IRA scores declined across most of the CT constructs, while the ICC substantially improved. This suggests that the changes made to the rubric helped to improve the tool's reliability overall, but clearer language in the descriptors was needed to better differentiate the levels and improve exact agreement among markers.

Based on faculty comments provided during and post blind marking, a number of factors were identified, which may have contributed to the low IRA results. These conditions included (but were not limited to):

- Mismatches between the assignments (what students were originally asked to demonstrate) and CT rubric requirements (what faculty blind markers were asked to assess);
- Lack of specificity to enable markers to distinguish between adjacent performance levels;
- Lack of definitions/descriptors to guide marking; and
- Students' poor writing skills that made it difficult for markers to assess CT skills.

Additionally, the use of numerical values (1–4) in the rubric added confusion. Faculty emphasized their concern about the rubric being based on four levels of performance when the college uses a grading system based on five letter grades. A decision was made to remove numerical values from the next iteration of the rubric. Faculty debriefing also identified a need to provide specific training in the use of the rubric for grading purposes.

Blind marking #3

Results from the third blind marking round indicate that the ICC scores declined but remained above target, except for two criteria. The IRA remained low and comparable to previous results. Based on faculty comments, the IRA likely continued to be affected by the same or similar problems encountered in the second blind marking. This group of markers also noted that they did not feel they had sufficient knowledge about the course or students in the same way the course instructors have in order to accurately assess the level of students' CT skills.

Table 1: Level of agreement and intra-class correlations (ICC consistency) for level of agreement and 95% confidence Interval (CI) for ICC											
Rubric 1-Blind Marking 1 (Feb-March 2014) (6 College English faculty)				Rubric 2-Blind Marking 2 (December 2014) (13 faculty; including 6 project participants)				Rubric 3-Blind Marking 3 (April-May 2015) (21 faculty; including 6 project participants)			
Criteria	IRA	ICC	95% CI	Criteria	IRA	ICC	95% CI	Criteria	IRA	ICC	95% CI
C1. Identification of issue/problem	0.5	0.58	-0.49-0.95	C1. Identification of issue. Take a position	0.53	0.91	0.74 - 0.99	C1. Clarifies the issue to be discussed and/or position to be argued in this paper	0.48	0.69	0.16 - 0.95
C2. Support for position taken	0.6	0.51	-0.71-0.94	C2. Choice of support sources	0.51	0.83	0.54 - 0.97	C2. Identifies the sources of ideas or evidence used in developing the argument or conclusions	0.52	0.91	0.75 - 0.99
C3. Language used in response	0.67	0.82	0.38-0.98	C3. Analysis of support	0.45	0.88	0.67 - 0.98	C3. Analyzes the ideas or evidence to develop the argument or conclusions	0.52	0.81	0.49 - 0.97
C4. Contradictory positions and underlying assumptions and biases	0.63	0.57	-0.5-0.95	C4. Contradictory evidence and expert's opinions and/or methodology	0.47	0.87	0.63 - 0.98	C4. Critiques contradictory evidence, information, expert's opinions and/or methodologies	0.49	0.73	0.10 - 0.98
C5. Conclusions	0.73	0.74	0.08-0.97	C5. Personal biases	0.54	0.9	0.72 - 0.98	C5. Acknowledges personal biases or assumptions	0.48	0.87	0.66 - 0.98
				C6. Conclusions	0.55	0.97	0.91 - 1.00	C6. Describes conclusions	0.53	0.85	0.59 - 0.98

Markers' Feedback About Their Experience Using the CT Rubric

Blind markers' feedback about their experience using the rubric in both blind marking rounds (versions 2 and 3) was positive overall, while noting some specific challenges and suggestions for improvement. Markers who participated in previous blind marking sessions agreed that the third version of the rubric was an improvement from the previous one. Finally, while the intent of the blind marking activity was to develop a valid tool to assess the acquisition of CT among George Brown College students, the process engaged faculty participants and volunteers in different ways, resulting in unintended positive outcomes, especially around changes in attitudes (acceptance/intention), behaviour (application of knowledge/new concepts), and self-reflective thought about one's teaching practice.

"I have made Critical Thinking assignments explicit in several of my courses to highlight them and make students aware of what they are learning." — *Faculty (marker) participant*

"As a result of this blind marking assignment, I am focusing so much more on the critical thinking skills you have raised in this rubric and, therefore, I will have a better idea of how/whether the rubric reflects the teachable skills at the end of this term!" — Faculty (marker) volunteer

"[The rubric] has left me wondering how I might change my teaching to address some of these skills (specifically, critiquing contradictory evidence)." — Faculty (marker) volunteer

"My first reaction as I read this latest iteration of the rubric was to say — aloud — 'Wow, I'm going to use this!' …. Though I am still half convinced this is actually an assessment of whether faculty assign grades based on skills in writing and presentation rather than as an objective response to the stated assignment criteria." — Faculty (marker) volunteer

The qualitative analysis can be found in Appendix E.

3.2 EES Curriculum Interventions

Given that the main objective of this project is to build capacity in assessing CT skills, ensuring that the desired skill is actually being taught and practised is also viewed as an integral component of this work; that is, identifying CT constructs and developing a rubric to assess CT skills alone is not sufficient without a review of current curriculum and teaching practices. The literature points to major commonalities in teaching critical thinking, including modeling critical thinking, finding problems and allowing students to practise critical thinking and giving feedback to students to allow them to improve (Brookfield, 2011; Hemming, 2000; Broadbear, 2003; Bensley, 2010). For the purpose of this project, starting in Phase 2 and in consultation with the curriculum specialist of the project team, the General Education/Liberal Arts faculty participants made three types of changes and revisions to their courses and teaching practice:

Course outline changes to reflect current practices (adjustments to EES matrix, learning outcomes):

- Each George Brown course outline includes an EES chart, which is meant to identify which EES are being addressed in the course. Some of the courses involved in this project were already teaching and evaluating CT skills but this was not identified in the EES chart on the course outline. The EES charts for all courses involved in this project were adjusted to identify teaching and evaluating of CT skills.

- Several course outlines included learning outcomes that did not reflect the deeper learning that was being infused into the course teaching and learning strategies because of involvement of the project. Outcomes that began with "Recognize" and "Understand" required further discussion and revision to reflect the CT activities that were being added to the classroom practice.

Assessment tool changes/modifications to incorporate the newly developed CT assessment rubric in full or partially:

- This is where the greatest changes were made because of faculty involvement in the project. All participating teachers looked at their major assignments through the lens of the revised CT rubric and made changes based on the CT criteria.

- Generally, even though some assignments were not changed significantly, there was a greater awareness of CT skills because of revisions to other assignments, and as assignments are changed over time, more of the CT constructs may be addressed.

- Examples:

    o A nutrition assignment that was initially describing a diet for a special-needs client added an evaluation component of the diet and recommendations for change.

    o A course that did not appear to have any assignments that included analysis of personal bias added an assignment that asked students to answer questions about their knowledge and opinions of the topic at the beginning of the course and then analyze, with references, the changes based on their learning in the course and add what changes they will make in their lifestyles because of this.

    o A geography assignment that added a component of recommendations based on research.

    o A globalization assignment that was initially a summary of a video or article and was changed to a comparison of ideas presented in more than one video or article.

Classroom practice changes — revision of assignments, teaching strategies, exercises:

- Several courses introduced active-learning strategies to replace lectures and/or added more activities that encouraged deeper learning.

- Examples:

    o A nutrition class where students were taught about all the vitamins, including recommended dosages, where to get them in food and how they contribute to the body. These concepts were taught using a jigsaw method where students became experts on their own vitamin, and then taught other experts and were taught in turn. Students were active and involved in the class.

    o A psychology class where student groups were given a particular concept, then had to argue for and against the concept, and present their arguments to the class as a whole.

    o A science class where student groups were given all the information required to determine the conditions in which coral reefs thrive and were asked to develop a list of conditions.

- Teachers acknowledged that they had not previously been in the habit of talking about or describing CT per se, but still expected to see CT skills demonstrated in students' graded work. As a result of this work, the teachers began to talk about and describe CT to their students more explicitly and to identify concrete examples of CT in classroom discussions and, more specifically, in assignments.

3.3   Faculty Participants' Reflection/Experience with CT and Project Overall

At the end of the project, the six faculty participants and the curriculum specialist facilitator were asked to reflect on their experience in developing and using the CT rubric and their involvement in the project.

Faculty identified that their own involvement in the project provided them with greater insight and clarity about the many constructs underlying CT, including the six constructs that were ultimately chosen to be included within the CT rubric, and about their own expectations and assumptions regarding students' CT knowledge and skills.

Faculty acknowledged the importance of identifying and addressing CT more deliberately and more explicitly with their students in order to increase students' general awareness of critical thinking and of the relevance and benefits of developing CT skills.

Faculty identified various opportunities to introduce and discuss CT constructs with their students during the semester, both in general terms and through the incorporation of CT constructs within specific assignments and the creation of new assignments that explicitly included CT tasks. This included general classroom discussions about CT and the integration of CT tasks and clearer expectations for assessing CT within their assignments. They noted improved quality and consistency of grading (i.e. less subjectivity, more clarity for both professor and student).

In using the CT rubric, faculty identified certain challenges. Some found that not all six CT constructs used in the rubric were relevant to their needs, for example the construct which involved identifying personal bias. Some faculty found it necessary to introduce other CT constructs that were felt to be more germane to their course. Some adapted the CT rubric in order to meet the grading criteria of their specific assignments (this flexibility was intended in the planning of the rubric).

The curriculum specialist also confirmed that helping faculty to make changes to their teaching practice and incorporating these CT constructs into actual curriculum (what is taught, practised, assessed) requires dedicated time and skillful facilitation.

More detail is provided in Appendix F.

3.4   Discussion/Summary

The overall objective of this component of the project was to address three research questions:

- Do our existing assessment tools allow us to measure critical thinking?
- What are the measurable elements of critical thinking relevant to a postsecondary education?

- What would be the elements of a generic assessment tool for critical thinking that could be adapted for use in any course in the college?

To answer these questions, the project team and faculty participants undertook several research activities, including a review of the existing assessment tool used by COMM faculty; several consultations leading to the identification of CT constructs; rubric development and validation; and changes to course design and delivery. The findings are summarized below.

Do our existing assessment tools allow us to measure CT?

The answer is no. Through the initial phase of this project, researchers learned from COMM faculty that CT is indeed being addressed and skills are indeed being taught within the classroom in a variety of ways. However, critical thinking per se is not explicit within the English college course learning outcomes, not specifically identified within the curriculum, not specifically assessed by COMM faculty in college English courses and, as a result, not reflected in the existing COMM rubrics.

What are the measurable elements of CT relevant to a postsecondary education?

Using insights from the literature and consultations with COMM faculty in Phase 1, the project team identified the first set of CT elements, consisting of five CT constructs and a communication checklist. The constructs were used to develop the first CT rubric, which was tested for reliability. In consultation with General Education/Liberal Arts and other faculty participants, two major revisions were made to the CT constructs and rubric between 2014 and 2015. Although the project team and faculty working group acknowledged that there are other CT constructs found in the literature, the team agreed on the final selection of six CT constructs identified by faculty to be most relevant to college level courses. These constructs are:

- Clarifies the issue to be discussed and/or position to be argued in this paper
- Identifies the sources of ideas or evidence used in developing the argument or conclusions
- Analyzes the ideas or evidence to develop the argument or conclusions
- Critiques contradictory evidence, information, expert's opinions and/or methodologies
- Acknowledges personal biases or assumptions
- Describes conclusions

Overall, the blind marking activities provided faculty and volunteers with an opportunity to play a key role in the validation and testing of a rubric that could be used by teachers across the college. While the revisions to the rubric made the tool reliable, future versions could improve exact agreement among markers by adding definitions, making further adjustments to the criteria and level descriptors, and ensuring that the assignment and rubric requirements fully match. During the course of the project, it became evident to the working group that a period of familiarization with the rubric among faculty and students along with greater emphasis on what constitutes CT will be required if the goal is to embed it across the curriculum. Overall,

one of the greatest benefits of this activity, is that it provided a vehicle for faculty to share and discuss ways in which they are embedding CT in their courses and the difficulties associated with doing so. In addition, some faculty participants and volunteers explicitly credited their participation in the blind marking rounds and/or their involvement with the project for encouraging self-reflection about CT and for giving them a new tool — the rubric — which they could use to make the language of CT more explicit to students in class and in course outlines and assignments.

What would be the elements of a generic assessment tool for critical thinking that could be adapted for use in any course in the college?

Our fundamental objective of this project was to "build capacity to measure EES," in this case CT skills. That is, we wanted to develop a flexible tool that would enable faculty teaching various subjects across the college to quickly identify the core CT constructs and be able to incorporate these constructs into existing or new assignments in a way that could be easily graded. At the same time, the project team acknowledged that the identification of CT elements alone would not be sufficient to fully build institutional capacity to measure those CT elements. As a result, with the guidance of the curriculum specialist from the project team, faculty participants engaged in three types of curriculum/teaching changes:

- **Course outline** changes to reflect practices (adjustments to EES matrix, learning outcomes)
- **Assessment tool** changes/modifications to incorporate the newly developed CT rubric
- **Classroom practice** changes — revision to assignments, teaching strategies, exercises

To provide a breadth of perspective, we engaged faculty whose teaching assignments included English and General Education and Liberal Studies courses across the college as well as the students enrolled in those General Education and Liberal Studies courses. Faculty participants identified a number of clear benefits through participating in discussions with their colleagues, in identifying and clarifying the CT constructs, in identifying and focusing on the six CT constructs judged to be most relevant to their teaching, in explicitly communicating and explaining these CT constructs to their students, and in using the new CT rubric in both the design and grading of assignments. As a result, faculty participants changed their teaching practices and increased the focus on CT within their classrooms and assignments, regardless of the subject taught.

Capacity to teach and assess critical thinking was, therefore, increased among these faculty.

As noted in the preceding section, despite having participated in its development, the same faculty also identified a number of challenges in actually using the new CT rubric in their own practice, which the college needs to consider for a college-wide implementation. Having identified these challenges, faculty still celebrated the CT rubric tool as an excellent starting place for discussion of critical thinking with their students and appreciated the flexibility of the CT rubric design, which allowed for adaptation to their specific needs. In this respect, the CT rubric provides an efficient tool for other faculty teaching any subject to incorporate CT constructs into their teaching practice and in designing student assignments.

Limitations of the CT rubric

While the CT rubric is still a work in progress, we have identified some limitations during the development process:

- The early blind marking methodology utilized written assignments that were not originally developed with and did not reflect the CT constructs used in the rubric. This contributed to weak reliability scores as noted above. Subsequent blind marking sessions utilized more relevant assignments and revised constructs and descriptors, which led to an improvement of inter-rater reliability, but inter-rater agreement remained below the benchmark. Future attempts at blind marking to validate the evolving CT rubric should use only assignments that were developed to specifically reflect all of the core CT constructs.
- We have learned that there is no single set of CT constructs that reflects all types of college assignments requiring critical thinking. Despite our intention to create an assessment tool that would be generic enough to be used for any subject taught, early users of the CT rubric felt that some constructs were not relevant to their assignments or were missing from the rubric. It is recommended that the college uses this insight to review, update and revalidate the CT rubric on a periodic basis to ensure it remains relevant and reliable.

4. Student Self-assessment Tool for CT Skills

Initially, it was thought that the validation of the CT assessment tool would be strengthened by triangulating markers' scores with an additional data set, that being students' assessments of their own critical thinking on a marked assignment given the same criteria as those used by the marker. It was also intended that a student self-assessment tool would help to answer the research question, "Are students aware of their own CT skills?" With this in mind, we set out to develop a student self-assessment tool for critical thinking that could eventually be used in conjunction with the CT assessment rubric as part of the validation process.

4.1 Method

The literature review identified limited research on the self-assessment of critical thinking. Studies looking at student self-assessment and peer assessment reflected a long-standing debate about the validity of self-assessment questionnaires, particularly in the absence of objective performance feedback. Studies such as Kruger and Dunning (1999) demonstrate that certain subgroups of students tended to inflate or underestimate their own grade, resulting in reliability issues. Others such as Ackerman et al. (2002) and Sarin and Headley (2002) have demonstrated a positive correlation between students' self-assessment and objective measures. Recommendations to strengthen the relationships between student self-assessment and objective assessments include periodic feedback on student performance throughout the course. As well, students should practise self-assessing and receive feedback on the accuracy of their self-assessments as it may be a skill that students can obtain through practice and feedback (Sitzmann, Ely, Brown, & Bauer,

2010). These recommendations, while noted by the project team to be important, were not included within the original scope of this project.

In her doctoral work, Nair (2011) also reviewed the many challenges inherent in the self-assessment of critical thinking. Ultimately, Nair published a comprehensive critical thinking self-assessment for nursing students, which included an extensive list of 115 statements reflecting and aligned with the CT conceptual framework described by Facione (1990). In designing the GBC student self-assessment tool, of Nair's original 115 self-assessment statements, 43 were selected, which the project team felt aligned best with the CT constructs included in the early version of the CT assessment tool. This list was further narrowed down to reduce conceptual duplications, then pilot tested and analyzed with a panel of student volunteers and revised accordingly. The final version of the student self-assessment tool included 13 statements, each about a specific element of critical thinking. Each first-person statement began "I know how to…" and was rated on a 7-point scale ranging from 0 (not at all knowledgeable) to 6 (extremely knowledgeable). The self-assessment survey also included an anonymous personal self-identifier (mother's maiden name and student's month and day of birth) that was used to link responses from the entry-exit surveys to that unique student identifier. Refer to Appendix G and Appendix H for the entry and exit questionnaires.

The self-assessment tool was administered to students at the start and end of the term in Phase 2 in courses in which critical thinking was not explicitly taught to provide a baseline measurement for comparison with the pre-post survey of students in Phase 3 in courses in which critical thinking was explicitly incorporated into the curriculum and highlighted as a learning objective. The goal was to identify whether there were significant differences in responses between the control and intervention groups about their self-assessment of CT skills.

4.2   Survey Administration

In Fall 2014, the self-assessment surveys were administered to students registered in six General Education/Liberal Arts courses, two each from Arts and Humanities, Science and Technology, and Social Sciences. The survey was conducted in class in Fall 2014 at the beginning of the term in weeks 1 and 2 and again in weeks 11 and 12. These courses were part of the control group in which CT skills were not explicitly taught. In Fall 2015, the same surveys were administered in the six courses taught by General Education/Liberal Arts faculty participating in the project (intervention group): two courses from Social Sciences, three from Science and Technology and one from Arts and Humanities. As a result of the uneven course selection, the distribution of responses in the intervention group was skewed toward Science and Technology. Additionally, the only course chosen from Arts and Humanities was delivered online while the others were delivered in class. This affected both the distribution of responses by study area as well as survey completions for the online course.

4.3   Results

Combined, 747 students completed the entry and/or exit surveys: 356 unique students in Fall 2014 (control group; 83% of course enrollment), and 391 students in Fall 2015 (intervention group; 77% of course enrollment). Not surprisingly, a comparative analysis of both groups of respondents indicated that the groups differed across the variables under study. Since the project's intent was to assess differences in students' self-assessment of CT over time and between control and intervention groups, it was felt that the analysis should focus on the subgroup of students who completed both surveys. Using students' unique identifiers, researchers were able to match responses from 77 students in Fall 2014 and 94 students in Fall 2015. The response rates of students who completed both surveys in relation to course enrollment were close to 20% in both study years (Table 2). The lower response rates for these groups of students may be explained by the fact that some students did not complete or provided different information during the entry-exit surveys necessary to create a truly unique identifier to match their responses.

The group of students who completed both surveys in Fall 2014 and Fall 2015 were similar in terms of age and prior educational attainment. Their response distribution differed by study area and study year. The Fall 2015 intervention group had a higher proportion of respondents from Science and Technology, from the Centre of Preparatory and Liberal Studies, and from first year compared to the Fall 2014 control group. Responses from this subgroup of students were analyzed to identify significant differences between and within groups. Results are presented in this section of the report. For further details, refer to Appendix I, Table 1.

Survey results for the entire group of respondents are available in Appendix I, Tables 6 and 7.

Table 2: Student self-assessment response rates									
Survey period	Course enrl.	# Resp. entry and/or exit surveys	Resp. %	# Resp. Entry-survey	Resp. % Entry-survey	# Resp. Exit-survey	Resp. % Exit-survey	# Resp. both surveys	Resp. % both surveys
Fall 2014 (control)	427	356	83.4%	267	62.5%	166	38.9%	77	18.0%
Fall 2015 (intervention)	508	391	77.0%	320	63.0%	165	32.5%	94	18.5%

Within Groups Analysis

A comparative analysis between the entry and exit surveys within the control and intervention groups revealed a very similar pattern: overall, students rated themselves relatively high in their ability to perform nearly all 13 CT skills at the start of the course and even higher at the end of the course. The differences were statistically significantly higher for six CT skills in the control group and for five in the intervention group; albeit the differences observed were for different aspects of CT, as shown in the table below.

Table 3: CT related skills rated statistically higher at the end of the course than at the start of the course (scale: 0-6)					
Control group (mean)	**Entry**	**Exit**	**Intervention group (mean)**	**Entry**	**Exit**
Given a situation or topic to consider, I know how to identify the main issue(s)	4.6	4.9	I know how to tell when additional information is needed in order for me to confirm or reject a given claim	4.4	4.7
I know how to tell if the evidence presented by the author is relevant to the argument	4.5	4.8	I know how to determine the similarities and differences among the various opinions presented for a given problem	4.7	4.9
I know how to systematically analyze the problem using multiple sources of information to draw conclusions	4.4	4.7	I know how to tell if the information supporting the argument is reliable	4.4	4.7
I know how to tell if the information supporting the argument is reliable	4.4	4.7	I know how to explain how my personal preferences/bias may affect my analysis	4.4	4.8
I know how to arrive at conclusions that are supported with strong evidence	4.6	4.9	I know how to develop logical conclusions	4.6	4.9
I know how to clearly explain my reasons for accepting or rejecting an author's claims	4.4	4.7			

Between Groups Analysis

A comparative analysis of responses between the Fall 2014 and Fall 2015 groups revealed the following:

- There were no significant differences in the way both student groups self-assessed their ability to perform CT related skills at the start of the course. Both groups rated themselves relative high, with mean scores ranging from 4.4 to 4.9 (on a scale from 0, not at all knowledgeable, to 6, extremely knowledgeable). Their responses aligned with their answers about prior instruction in CT. In the control group, 95% of respondents and 94% of those in the intervention group

reported some training (67% –70%) or substantial training in CT (28%–23%) prior to this course. See Appendix I, Table 2.

- When the information is broken down by level of education attainment, responses from the control group appear to be positively correlated; that is, the higher the level of education, the higher the proportion of respondents who indicated receiving substantial instruction/training in CT. This was not the case for the intervention group. Those with high school or less were as likely to report substantial instruction/training in CT as students who completed PSE. See Table 4.

- Overall, both groups completed PSE prior to entering their programs at similar rates (control, 25%; intervention, 23%). However, a higher proportion of students in the control group (42%) reported receiving substantial instruction/training than students in the intervention group (23%). See Table 4.

		Table 4: Prior CT instruction/training by highest level of education		
		How much instruction or training have you received previously related to the above [13 CT statements]…?		
		I haven't received instructions/training	I have received some instructions/training	I have received substantial instructions/training
Fall 2014	High School or less (n=45)	4.4%	75.6%	20.0%
	Some PSE (n=11)	0.0%	63.6%	36.4%
	PSE (n=19)	10.5%	47.4%	42.1%
	Total (n=75)	**5.3%**	**66.7%**	**28.0%**
Fall 2015	High School or less (n=55)	9.1%	65.5%	25.5%
	Some PSE (n=17)	5.9%	76.5%	17.6%
	PSE (n=22)	0.0%	77.3%	22.7%
	Total (n=94)	**6.4%**	**70.2%**	**23.4%**

- Both groups rated themselves similarly (higher) at the end of the course, suggesting that students in the control group reported knowledge gains in CT at the same rate as those from the intervention group. See Appendix I, Table 3.
- No significant differences in responses were observed for the majority of the 13 CT related skills by age and prior educational attainment for both student groups in either exit or entry surveys.
- To the question on whether students felt their CT skills had improved as a result of the course, a higher proportion of students in the intervention group reported more positive changes than the control group overall; albeit the differences are not statistically significant. See Table 5.

Table 5: CT skills improvement over the term			
	Fall 2015 (n = 75)	Fall 2016 (n = 94)	p-value
Improvement in CT			
Substantially improved	7 (9.3%)	14 (14.9%)	
Somewhat improved	44 (58.7%)	62 (66.0%)	0.236
No change	17 (22.7%)	12 (12.8%)	
Don't know	7 (9.3%)	6 (6.4%)	

Differences in proportion are compared by Pearson Chi-square test.

Definition of Critical Thinking: The Student Perspective

As part of the follow-up survey, students were asked to define critical thinking during the exit survey. The goal was to assess whether students from the intervention group, in which CT was overtly identified and discussed, would be better able to articulate and/or identify the constructs of CT than students from the control group.

The following provides findings based on input provided by students who completed both entry/exit surveys in Fall 2014 (control) and Fall 2015 (intervention).

Method

Of the 77 students who responded to both surveys in Fall 2014, 66 provided a definition of CT, and of the 94 students who completed the survey in Fall 2015, 89 provided a definition.

The majority of respondents in the Fall 2014 survey (67%) were from year 2 and above. In Fall 2015, there was a better representation of students by study year: 44% were from year 1 while the remaining 56% were from year 2. The last version of the CT rubric (#3) was used as a guide for coding responses.

Analysis

The majority of respondents identified at least one aspect of GBC's CT constructs. Students were able to articulate a wide range of CT skills, from basic concepts to more elaborate ones that included different CT skills:

"Being able to process information and come to a conclusion." — Year 1 Fall 2014 student

"Using available evidence and considering multiple perspectives to draw out a logical, valid answer or conclusion that is unbiased." — Year 1 Fall 2015 student

Fewer than 10% of respondents provided very vague or redundant answers that may indicate limited knowledge of CT or unwillingness to expand on ideas, among other things.

Students' definitions of CT as described in rubric #3 were limited to one construct in Fall 2014 and two constructs in Fall 2015. In Fall 2014, the most commonly cited definitions covered components related to "analyzes the ideas …" (25% of mentions) and to a lesser extent to "describes conclusions" (11% of mentions). "Clarifies the issues to be discussed" and "acknowledge personal bias" received very few mentions. Although bias was not mentioned per se, few students noted that being objective and able to consider all aspects of an issue are important aspects of CT. In Fall 2015, students' definitions covered components of "analyzes the ideas"' and "critiques contradictory evidence" equally (19% of mentions each). "Clarifies the issue"; "identifies sources"; "acknowledges personal bias"; and "describes conclusions" received few mentions.

CT in relation to its application and use were top of mind among students. It is interesting to note that students' definitions incorporated aspects of CT in relation to its application and use in different contexts, which serve to highlight other elements of CT that GBC may want to consider. This theme received 20% of mentions in Fall 2014 and 14% of mentions in Fall 2015.

"Taking the knowledge that you have learned (in and out of school) and applying it to other situations or to problem solve." — Year 2 student Fall 2014

"[CT] is the ability to analyze data and then apply that analysis to solve problems or come up with creative solutions." — Year 2 student Fall 2014

"The ability or application of an ability that allows a person to use the information given to solve a problem, normally causing the person to have to utilize the data to its fullest extent." — Year 2 student Fall 2015

4.4   Discussion/Summary

It was intended that this self-assessment tool would help to answer the research question, "Are students aware of their own CT skills?" The entry and exit course self-assessment surveys of CT provided some valuable information about students' self-perceived ability to perform various CT related skills. The comparative analysis within the control and intervention groups indicates that students in both groups feel relatively competent in their ability to perform CT tasks at the start of the course and even more so at the end of the course. Researchers were specifically interested to find any significant differences between the control and intervention group in their self-assessment of performing 13 CT skills, and found no differences overall. However, while both groups credited their General Education/Liberal Arts courses for enhancing their CT skills over the course of the semester, the intervention group acknowledged higher improvements in CT as a result of the General Education/Liberal Arts course. It is hypothesized that this may be due to the fact that faculty participants overtly discussed CT in their classroom. Based on students' definitions of CT, it was confirmed that the majority of respondents appear to have a general understanding of what constitutes CT as defined by GBC; albeit only a few constructs of CT, especially "analysis," received significantly more mentions than other constructs. Furthermore, some students went further by identifying other constructs of CT that are found in the literature but are absent from our CT rubric.

To the question, "Are students aware of their own CT skills?" findings indicate that students appear to believe so. However, what is still unknown is what level of mastery these students are at and whether their self-assessment of CT skills would match their actual demonstration of these skills, if tested. This subject is beyond the scope of this project.

Limitations

A limitation of the student self-assessment survey is that it was created in 2014 while consultations with GBC faculty were still underway to identify CT constructs relevant to GBC for the creation of the CT rubric. Therefore, the student self-assessment survey covers constructs found in the literature, but it does not fully align with those constructs ultimately used in the CT rubric. Secondly, the self-assessment survey is an indirect measure of CT skills, as self-reported by students and unrelated to specific work. The inclusion of an objective assessment of students' CT skills at the start and end of the term, or linked to a specific assignment, would have provided far richer data.

Given what we have learned thus far about student self-assessment of CT, we determined that further work was required to align the student self-assessment with the constructs used in the CT rubric and to test out the use of both tools on specific marked assignments. Therefore, student self-assessment was deemed not yet ready to be used as part of the process of validation of the CT rubric, as was originally planned. Future work would include revision of the student self-assessment and implementing a study in which student-specific responses to the self-assessment would be tied to a specific assignment and then triangulated with the marker's objective score of the student's work using the CT rubric.

5. Knowledge Transfer

One of the key objectives of this project is to develop tools that can be adapted and applied by faculty within any program curriculum across the college and, ultimately, within any postsecondary curriculum. And in doing so, the ultimate goal is to build faculty capacity to measure EES attainment.

This project specifically enabled the college to pilot a highly collaborative process for developing a valid assessment tool for one particular EES, critical thinking. With an expert curriculum specialist available to mentor General Education/Liberal Arts faculty participants to enhance their skills in teaching and assessing CT skills, insights gathered from the project led to the development of the *Critical Thinking: Learning, Teaching and Assessment — A Teacher's Handbook,* which included both the new CT rubric as well as resources to explain and support the use of the six CT constructs in the development of curriculum and assignments. The handbook was distributed electronically to all full- and part-time faculty at George Brown College in September 2015 and presented in four faculty workshops. The handbook is posted as a downloadable teaching resource on the college's internal website and will also be available through HEQCO's online Learning Outcomes Assessment Resource Room. Hard copies of this handbook have been distributed widely within the college and at two scholarly conferences.

This project team is committed to knowledge transfer and dissemination of findings through workshops and scholarly presentations both within and outside the college. The handbook and project have thus far been presented at two scholarly conferences: the Curriculum Developers Affinity Group annual conference (November, 2015; London, Ontario) and the CIDER Conference on Higher Education Pedagogy (February, 2016; Blacksburg, Virginia).

The project team has also completed *Essential Employability Skills — A Teacher's Guide*. This companion guide is a practical teachers' handbook encompassing all six categories of the essential employability skills. This guide will help teachers integrate EES into their curriculum with practical how-to suggestions and examples demonstrating how learning, practising and evaluating EES might look in the classroom. This document will also be available to the public.

The six faculty participants, with their curriculum specialist mentor, will continue to explore their own teaching practices and use of the CT constructs and rubric within the courses they teach, to build awareness of their work among colleagues, and to offer support to other faculty at George Brown College in focusing on critical thinking in their teaching and assessment materials.

6.   Conclusion/Recommendations

The ultimate objective of this project is to build capacity among faculty to measure essential employability skills. This project enabled the college to focus on one of the six categories of EES, critical thinking, and to pilot a collaborative process for doing so.

Over the past three years, the project team engaged numerous faculty in the development and piloting of practical new curriculum resources and an assessment tool for the learning, practice and demonstration of critical thinking within a course. These resources were successfully piloted within a selection of General Education/Liberal Arts courses, reflecting the broader intention that these resources could be adapted and applied by faculty within any program curriculum across the college and, ultimately, within any postsecondary curriculum.

The project team believes that the development process piloted during this project can be replicated with other EES, given that the resources are available to support the initiative. The project has also provided a number of key insights and lessons learned that should be considered in followup to this project and for other institutions considering undertaking similar work.

The Importance of Faculty Engagement

Faculty engagement in the identification of core CT constructs to be taught, practised and assessed was crucial. This provided a broad range of perspectives and interpretations of the CT constructs, identified a range of curriculum contexts in which critical thinking could be taught and demonstrated, and helped to increase faculty acceptance of the end products. Guided by the CT literature and mentored by a dedicated curriculum specialist, our faculty volunteers collaborated to identify the CT constructs judged to be most relevant to their unique curriculum and college-level assignments. Judging from the positive feedback received from both project faculty volunteers and other faculty to whom the tools have been presented, we believe the collaborative and mentored development process was an essential element of achieving success in creating end products that can be applied in a broad range of contexts. This in turn should enable students to acquire transferable CT skills that can be applied in any context as posited by the literature.

Challenges in Developing a Generic Assessment Tool

We started the project with the philosophy that critical thinking could best be assessed in a relevant context in which it is taught and practised, as opposed to employing a generic standardized test. We learned that critical thinking was not explicitly being addressed or assessed within existing college English courses and rubrics. We learned that assessing critical thinking was complicated or obscured by the quality of students' English language skills and that faculty required additional training to separate out assessment of critical thinking and language. We learned that faculty had many different interpretations of critical thinking and how it could be demonstrated in student assignments within their unique contexts and that considerable discussion was required to arrive at a consensus about the specific constructs to be included in a generic

tool and how to best describe generic performance levels for each. Our blind marking rounds demonstrated the challenges of finding appropriate and relevant samples of student work to test the evolving rubric when available student samples were based on previous assignments that did not explicitly incorporate the constructs being assessed by the rubric.

After multiple revisions of the CT constructs, corresponding rubric and testing, the inter-rater reliability scores (ICC) of the rubric improved considerably, in most cases well above the set benchmark, making the rubric, as a tool for assessing various CT constructs, reliable overall. However, we have further work to do to improve inter-rater agreement, especially for adjacent performance levels.

Assumptions About What Students Know

Our faculty volunteers acknowledged at the outset that they did not necessarily address critical thinking per se in their courses, even if it was identified in the EES chart on the course outline (that is, they did not talk about CT, did not explicitly demonstrate CT or enable students to practise CT, and did not identify CT as something that was specifically being assessed in assignments). In doing so, they were making untested assumptions that students had already learned about CT elsewhere and had developed some skills in this area that could be applied in their courses. They were also making assumptions that students understood that they were being taught CT skills without making that explicit and expected students to be able to demonstrate those skills in their assessments. The same could be said of the other 10 mandatory EES, which are not routinely incorporated into the college curriculum.

We hypothesized that if CT was more explicitly addressed within the curriculum (discussed, learned, practised, evaluated), then students would have both greater awareness and understanding of CT and would demonstrate improved CT skills in the course.

In our initial efforts to find out what students knew about CT and to demonstrate students' self-assessment of their own CT skills, we learned that students did actually possess some preliminary knowledge of critical thinking (in fact, students identified most of the constructs incorporated in the CT rubric and suggested several other practical constructs that should be considered in future), that students self-assessed themselves as relatively proficient at CT at the start of the course and even higher at the end of the course, and that students credited their General Education/Liberal Arts courses for enhancing their CT skills.

Similar to other discussions about student self-assessment in the literature, a significant limitation of this work is validity of self-assessment and the lack of an objective measure of each student's CT skills, taken at the start and end of a course, to assess the validity of these self-assessments. These should be considered in followup work.

Evaluating Only What has Been Taught and/or Practised

Further to the above, our faculty volunteers acknowledged that prior to the project, they had been assessing some elements of critical thinking without being explicit about this in their teaching or in the assignments being marked. This became evident when we sought relevant student samples for blind marking rounds. This was also demonstrated in the EES charts on course outlines, in which CT might or might not be checked off by the professor under the three headings: Learned, Practised or Evaluated. Again, the same might be claimed about the other 10 EES.

The facilitated curriculum changes that were a core element of Phase 2 of the project (what we labeled the "EES Curriculum Intervention"), which sought to ensure that CT was explicitly addressed by the professor and incorporated into the curriculum (identified up front in the course outline, specifically identified and discussed in class, incorporated into assignments and identified in marking rubrics).

Faculty feedback (from the project volunteers and others) confirmed that developing an assessment tool on its own provided no benefit unless all the elements to be assessed were also clearly incorporated into the curriculum. Through this project, we have identified a collaborative, inclusive process and the pedagogical resources needed to meet that objective.

Conclusion

In conclusion, we believe this project was successful in meeting our original objectives:

- To develop and pilot test practical and relevant new curriculum resources and a validated assessment tool for the learning, practice and demonstration of the essential employability skill of critical thinking within a course.
- To design these resources to enable adaptation and application by faculty within any program curriculum across the college and, ultimately, within any postsecondary curriculum.
- To pilot a collaborative developmental process for one EES that can be applied to the development of similar tools and resources for other EES.

This project has resulted in the following products:

A CT Assessment Rubric

We identified a core set of CT constructs, rooted in the literature, that are viewed as most relevant by college faculty. Based on these, we developed a reliable CT rubric that provides the college with an excellent starting point from which to assess students' CT skills consistently across the curriculum. Improving the reliability of the rubric is a work in progress.

Critical Thinking: Learning, Teaching and Assessment – A Teacher's Handbook

With the assistance of a curriculum specialist, a faculty working group incorporated the core set of CT constructs into their existing course curriculum, impacting on their course outlines, what was discussed and taught, and how this was assessed in assignments. Along with the CT rubric, practical suggestions and real examples for incorporating the CT constructs and how to use the rubric, taken directly from the work of this faculty group, were fashioned into a valuable handbook for teachers, and distributed widely. Faculty were encouraged to use this resource in their own classroom.

Essential Employability Skills: A Teacher's Guide

We created an additional teacher resource with practical examples and suggestions for incorporating all 11 EES into any college curriculum. This guidebook does not, however, contain rubrics for each EES.

A Pilot-tested, Collaborative Process for Developing Measurement Tools and Associated Curriculum Resources that can be Utilized for Other EES

The project was designed to enable a collaborative, inclusive developmental process to unfold, engaging a representative group of faculty volunteers and fostering significant change in curriculum and teaching practices and resulting in the products described above.

Those involved in the project believe the collaborative process was of great benefit in raising their and their students' awareness of CT, in changing their own teaching practices and in incorporating CT more explicitly within assignments.

Key to our success was the assignment of a dedicated curriculum specialist to facilitate discussions and work directly with faculty to help them identify opportunities to change teaching practice and to incorporate the relevant constructs into their curriculum (what is discussed, what is taught, what is practised by students, what is assessed). Such a professional development process, while effective and well received by faculty, is resource intensive and requires an ongoing commitment of resources in order to scale up across the college and be sustainable over time.

George Brown College will continue to build on the existing work and consider the lessons learned and recommendations provided by the project team.

References

Ackerman, P. L., Beier, M. E. & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences*, *33*(4), 587–605.

Allen, S. & Knight, J. (2009). A method for collaboratively developing and validating a rubric. *International Journal for the Scholarship of Teaching and Learning*, *3*(2), 1–17.

American Association of College and Universities (n.d.). *Critical thinking value rubric.* Retrieved from http://assessment.aas.duke.edu/documents/CriticalThinking.pdf

Bensley, A. D. (2010). A brief guide for teaching and assessing critical thinking in psychology. *Observer*, 10. Retrieved from http://www.psychologicalscience.org/index.php/publications/observer/2010/december-10/a-brief-guide-for-teaching-and-assessing-critical-thinking-in-psychology.html

Bers, T. (2005). Assessing critical thinking in community colleges. *New Directions for Community Colleges, 2005*(130), 15–25.

Bok, D. (2006). *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More*, Princeton, NJ: Princeton University Press.

Broadbear, J. T. (2003). Essential elements of lessons designed to promote critical thinking. *Journal of Scholarship of Teaching and Learning, 3*(3), 1–8.

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

Brookfield, S. (2011). *Teaching for critical thinking: Tools and techniques to help students question their assumptions.* San Francisco, CA: John Wiley and Sons.

Darrow, A. A., Johnson, C. M., Miller, A. M. & Williamson, P. (2002). Can students accurately assess themselves? Predictive validity of student self-reports. *Applications of Research in Music Education*, *20*(2), 8–12.

Facione, P. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. American Philosophical Association Delphi Research Report. Millbrae, CA: The California Academic Press.

Falichov, N. & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Education*, *59*(4), 395–430.

Gisev N., Pharm B., Bell J. S., & Chen T. (2013). Inter-rater agreement and inter-rater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy, 9*(3)*,* 330–338.

Graham M., Milanowski A., & Miller J. (2012). Measuring and promoting inter-rater agreement of teacher and principal performance ratings. Centre for Educator Compensation Reform.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449–455.

Hemming, H. E. (2000). Encouraging critical thinking: "But…what does that mean?" *McGill Journal of Education, 35*(2), 173–186.

Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130–144.

Knight, J. & Allen, S. (2012). Applying the PDCA cycle to the complex task of teaching and assessing public relations writing. *International Journal of Higher Education, 1*(2), p67–p83.

Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134.

Ku, K. Y. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking and Creativity*, *4*(1)*,* 70–76.

Kurfiss, J. G. (1988). *Critical thinking: Theory, research, practice, and possibilities*. *ASHE-ERIC. Higher Education Report No. 2*. College Station, TX: Association for the Study of Higher Education.

Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model*. Alexandria, VA: Association for Supervision and Curriculum Development.

Muñoz A. & Alvarez M. E. (2007). Students' objectivity and perception of self-assessment. *Journal of Asia TEFL, 4*(2), 1–25.

Nair, G. G. (2011) Preliminary Psychometric Characteristics of the Critical Thinking Self-Assessment Scale. Doctoral thesis. University of Saskatchewan, Saskatoon.

Paul R. & Elder L. (2007). Consequential validity: using assessment to drive instruction – White Paper. *Foundation for Critical Thinking*. Retrieved from http://www.criticalthinking.org/files/White%20PaperAssessmentSept2007.pdf

Paul R. & Nosich G. (n.d.). A model for the National Assessment of Higher Order Thinking. Retrieved from http://www.criticalthinking.org/pages/a-model-for-the-national-assessment-of-higher-order-thinking/591

Raupach T., Münscher C., Beibbarth T., Burckhardt G. & Pukrop T. (2011). Towards outcome-based programme evaluation: Using student comparative self-assessments to determine teaching effectiveness. *Medical Teacher, 33*(8), e446–e453.

Rickles, M. L., Schneider, R. Z., Slusser, S. R., WIlliams, D. M., & Zipp, J. F. (2013). Assessing change in student critical thinking for introduction to sociology classes. *Teaching Sociology, 41*(3), 271–281.

Sadler, P. M. & Good, E. (2006). The impact of self- and peer grading on student learning. *Educational Assessment*, *11*(1), 1–31.

Sarin S. & Headley D. (2002, June). Validity of student self-assessment. Proceedings of the *American Society for Engineering Education* Annual Conference and Exposition, Montreal, Quebec, Canada*.* Retrieved from https://peer.asee.org/validity-of-student-self-assessments.pdf

Sitzmann, T., Ely, K., Brown, K. G. and Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, *9*(2), 169–191.

Taylor, B. (2004). *Debating moral education—Topic 3: Teaching and learning.* Paper presented at Center for Academic Integrity conference, Duke University.

Wolf K. and Stevens E. (2007). The role of rubrics in advancing and assessing student learning. *The Journal of Effective Teaching*, *7*(1), 3–14.