Higher Education Quality Council of Ontario

An agency of the Government of Ontario

Assessing Learning Outcomes: Thinking Critically about Critical Thinking and Written Communication Skills

Vera Beletzan, Melissa Gabler, Paula Gouveia
Humber College

Published by

The Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402
Toronto, ON Canada, M5E 1E5

Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca
E-mail: info@heqco.ca

Cite this publication in the following format:

Beletzan, V., Gabler, M., Gouveia, P. (2017) *Assessing Learning Outcomes: Thinking Critically about Critical Thinking and Written Communication Skills.* Toronto: Higher Education Quality Council of Ontario.

Acknowledgements

The research team would like to acknowledge the countless faculty and staff at Humber College who volunteered their time to assist in the development and refinement of the critical thinking and written communications scorecard, piloted the tool in their classes and contributed to the overall success of this project. Their feedback and support over the entire project was critical to its success and truly enabled Humber to develop a scalable tool aimed at assessing learner gains in critical thinking and written communication, across programs, campuses and time. In particular, the authors wish to acknowledge the hard work of research assistant Sarah Cook and program assistant Carol Bueglas in the data collection during phase two of this project. Their dedication and commitment allowed the research team to work through a variety of constructs and contexts and their thoroughness and attention to detail strengthened the research study at each turn.

1. Executive Summary

Outcomes-based education (OBE), namely the emphasis in education systems on learning outcomes and their assessment, has had one of the largest and most significant impacts on postsecondary education (PSE) in recent decades. Not only does OBE present clear statements to describe students' skills and abilities, it also provides the vehicle by which postsecondary institutions can assess and improve the quality of their programs and demonstrate the value of these programs to both employers and the general public.

Ontario PSE institutions, colleges in particular, have long embraced OBE. From the development of postsecondary program standards that specify the vocational learning outcomes of their credentials, to the inclusion of general education requirements, and the introduction of outcomes that require graduates to demonstrate skills in communication, numeracy, critical thinking and problem solving, information management, interpersonal skills, and personal skills — known collectively as the essential employability skills (EES) — Ontario colleges have provided leadership in the development of OBE.

However, this leadership is not as uniformly evident when we consider the assessment of learning outcomes. Whereas numerous assessment approaches, as well as research to support their validity as measures of student performance, have evolved at the vocational level, in the case of the essential skills, and primarily of critical thinking (CT) — the focus of this study — there is to date no definitive assessment strategy. This situation is compounded by an ongoing lack of common understanding and consensus of what constitutes an essential skill such as CT, and the abilities that demonstrate its attainment.

The Higher Education Quality Council of Ontario (HEQCO) has long been interested in learning outcomes assessment both at the vocational and essential skills levels. In 2012, to further expand its research into this area, HEQCO led the development of a Learning Outcomes Assessment Consortium (LOAC) of which Humber College is a member. As part of this consortium, Humber College committed to developing a learning outcomes assessment tool to track students' development in critical thinking and written communication (CTWC) skills. This was completed as part of the first phase of the project (2013–2014). The college also committed to examining the feasibility and value of the tool with the goal of implementing it across the institution. This second phase required that we ask several research questions:

- What is the validity of the assessment tool?
- What is the reliability of the tool when it is used repeatedly?
- Are students making gains in CT and WC learning outcomes over time?
- Where do students experience greater gains in CT and WC?
- What are the usability and scalability of the tool?

From the fall of 2014 to the winter of 2016, 650 students from three of the college's eight schools had their written work evaluated using the assessment tool (scorecard) that was developed, piloted and revised during the first phase of the project. It is important to note that diverse student CT and WC learning experiences occurring in different learning contexts were selected for comparison: students' gains in CTWC

skills attainment were compared in cross-college reading and writing courses where these skills are explicitly taught and reinforced as course learning outcomes; in courses within a vocational program where these skills are implicitly embedded as they are deemed essential by the industry they serve; and also in a course where CT is taught as content knowledge.

Although conventional validity measures of any assessment instruments are challenging to establish, we generally found the assessment tool accurately captured the total CTWC skills scores, and that it had substantial content validity, concurrent validity and convergent validity. We also found the tool to be reliable in terms of its internal consistency when used repeatedly, as demonstrated by the consistency of faculty's ratings of the total skills scores on CTWC. At the same time, instructors' judgments on some of the individual CTWC skills categories of the scorecard were less consistent.

Upon analysis of the data, we did not find that students are making significant gains in learning the outcomes over the duration of the two examined courses in which CTWC skills are explicitly taught. Comparing levels of skill achievement among a small panel of students over only two periods of time likely contributed to the findings of minimal gains in learning. However, we did find students are experiencing greater gains in CT and WC in the different learning contexts that were selected for cross-sectional comparison. Students in those courses where CTWC skills are taught explicitly demonstrated the highest gains. Students in programs where CTWC skills are implicitly embedded in content showed lower levels of achievement than students in explicit learning contexts, but higher than students in discrete courses which focus on CT as a discipline (content knowledge). Overall, the findings suggest that, to achieve the highest gains, CT needs to be taught explicitly and as a skill. Moreover, this skill needs to be taught consistently and built over longer periods of time, which suggests that courses where these skills are taught explicitly should be positioned strategically throughout each program of study, ensuring maximum exposure and ample practice time.

Lastly, we examined the questions of usability and scalability. For successful implementation of the assessment tool across the institution, faculty need to find it easy to use. Furthermore, faculty need to consider the tool meaningful in terms of providing students with useful information regarding their level of CTWC skills, and how these skills can be improved. Given the significant degree of faculty involvement in the development, piloting and revision of the assessment tool, it is not surprising that almost three-quarters (74%) of the total sample of 46 faculty who used the tool agreed that it was easy to use. When examining faculty perceptions of the usability of the assessment tool by school, those faculty from the school that is primarily responsible for explicitly teaching CTWC skills had, on average, higher agreement than those from schools where the skills are only implicitly embedded. This suggests that faculty who explicitly teach CTWC skill building are both more familiar with the metacognitive skills at the core of CT and more comfortable with the assessment of these skills. On the other hand, a significantly lower percentage of faculty (58%) agreed that the assessment tool would provide meaningful information to students about their skill level. Additionally, opinions were more varied about its meaningfulness.

Given the results of the data analysis, we suggest that the assessment tool can readily be implemented across the institution but is best administered in courses where CTWC skills are taught explicitly.

Furthermore, there must be a significant focus on faculty training in order to improve overall reliability and, by extension, accuracy and validity. In addition, to increase the meaningfulness of the tool, not only must students understand which skills are being assessed and for what reasons, they must also have their skills assessed over a longer period of time with ample opportunity for them to make significant gains in learning, along with access to resources to allow for this. Finally, an institution-wide common understanding of the value of assessing essential skills beyond the obvious need for program accountability and quality is a prerequisite for developing, implementing and supporting both a user-friendly and meaningful CT assessment tool. This requires all stakeholders to think critically about CT and WC as essential skills that can and should be formally assessed as part of the learning outcomes of students in PSE programs.

Table of Contents

2. List of Tables

3. List of Figures

4. Introduction

For several decades now, postsecondary education (PSE) institutions in Ontario have characterized themselves as outcomes-based education (OBE) institutions. The focus on OBE has been reinforced by the introduction of credential frameworks by the Ministry of Advanced Education and Skills Development (MAESD 2009b), vocational program standards (MAESD, 2003), and discipline, or rather course-level learning outcomes. These last being clear statements of what students must know and be able to demonstrate through a variety of assessments, both formative and summative, upon completion of an individual course or program of study leading to a recognized credential (Kenny, 2011). Lennon and Jonker (2014) further posit that the focus on learning outcomes has provided postsecondary institutions with a means not only to measure student learning, but also to assess the quality of the education provided to learners and the value of this education to employers and the general public that so heavily invests in the public education system.[1]

While the value of OBE is no longer in question, education institutions still face the challenge of assessing these learning outcomes. From the introduction of the Assessment of Higher Education Learning Outcomes (AHELO) by the Organization of Economic Co-operation and Development (OECD) in 2006, to the implementation of "Tuning" projects to articulate learning outcomes in specific disciplines and credential levels, and the design of standardized tests such as the Collegiate Learning Assessment (CLA), the assessment of learning outcomes has gained considerable attention as PSE institutions focus on this type of assessment as a measure of accountability and quality. In fact, learning outcomes have been recently defined as "[c]lear indications of teaching and learning quality [that] are beneficial in a number of ways. They support better understanding of educational value to students, employers and the public at large" (Lennon & Jonker, 2014, p. 4).

The Higher Education Quality Council of Ontario (HEQCO) is actively involved in researching learning outcomes. HEQCO has led Ontario's involvement in the AHELO project, specifically in the civil engineering strand, provided ongoing support to university and college faculty in the design and assessment of learning outcomes in various discipline areas, and examined the feasibility of introducing standardized tools like the CLA. Currently, HEQCO continues to engage in meaningful work ranging from the development of learning outcomes to their measurement at both the local and international levels (Lennon, 2014; Lennon & Jonker, 2014).

In 2013, to continue the focus on addressing the challenge of assessing learning outcomes, HEQCO brought together a group of Ontario colleges and universities: Durham College, Humber College, George Brown College, Queen's University, the University of Guelph, and the University of Toronto, to form the Learning Outcomes Assessment Consortium (LOAC). This group was committed, first, to the development of effective

1 Lennon and Jonker (2014) also include increased international competitiveness as a benefit of the program improvements that result from a focus on learning outcomes. This same sentiment is echoed by Weingarten (2014).

learning outcomes assessment tools and techniques and, second, to their widespread implementation at their institutions. (Confederation College has since joined the consortium.)

As a member of the consortium, Humber College chose to focus not on the assessment of program or vocational learning outcomes, but rather on learning outcomes common to every college student's education experience, namely, the essential employability skills (EES), and, more specifically, on critical thinking (CT) and written communication (WC) skills. Our response has taken place over two distinct phases. The first phase, completed in 2013–2014, focused on the development of a tool to assess students' levels of achievement in CT and WC that could eventually be used and scaled successfully across the institution. Scorecard development, testing and redesign were accomplished in this pilot phase of the study. The second phase (2014–2016), which is the focus of this report, aimed to test the validity and reliability of the tool developed during the first phase. It also focused on assessing students' CT and WC skills over time and in the following three types of courses:

- Courses where these skills are explicitly taught and reinforced as course learning outcomes
- Courses within programs where these skills are more implicitly embedded in content because they are deemed essential by the industry they serve
- A course where CT is the actual focus of the course content

5. Research Objectives and Research Questions

The primary objective of this research project was to determine if the assessment tool — the Critical Thinking and Written Communication (CTWC) skills scorecard — is a reliable and valid instrument with which to track student progress in those specific essential skills. To respond to this primary objective, the research team focused its study on two central themes: validity and reliability, around which a series of detailed questions were asked.

Validity

- Does the CTWC skills scorecard measure what it is intended to measure?
- Do the individual indicators of the skills scorecard devised to measure CTWC really measure its components (construct validity)?
- Do experts agree that the individual indicators and overall measures reflect the concepts? Do faculty in the field agree (content and face validity)?
- How well do students' CTWC skills scores compare with other established measures (criterion validity)? For example, do students' CTWC skills scores correlate with their scores on an established CT skills measure such as the Test of Everyday Reasoning (TER) (concurrent validity)?
- Do students' skills scores correlate with other theoretically similar constructs such as grades on assessed work and final course grades (convergent validity)?

Reliability

- What is the consistency of the CTWC skills scorecard when it is used repeatedly?
- Is the scoring among faculty consistent (inter-rater reliability)?
- Do the individual items that make up the scorecard measure the same underlying construct consistently (internal consistency)?

The second objective of the study was to use the revised scorecard to track any progress in learning outcomes achievement among a cross-college diploma and advanced diploma program panel group in two reading and writing courses, given their early exposure to a special CTWC skills-building curriculum. To meet this goal, faculty used the scorecard over the span of the 2014–2015 academic year to assess select students' work in the panel group for any gains in learning the essential skills in two courses (WRIT 100: College Reading and Writing Skills and WRIT 200: Workplace Writing Skills).

A third objective was to offer some initial insight into whether diverse student CTWC learning experiences occurring in different learning contexts (i.e., the three types of courses listed in the previous section) have had any effect on student gains in learning CTWC essential skills outcomes. With respect to this objective, faculty applied the scorecard to select student work throughout the academic years 2014–2015 and 2015–2016 in the wider cross-college context (WRIT 100 and WRIT 200) as well as in the diploma-level Police Foundations Program (PFP 200: Criminal and Civil Law and PFP 211: Interviewing and Investigations) and certificate-level General Arts and Science (GAS) College and University Transfer Program (GCRT 100: Critical Thinking). To achieve these aims, the research team asked the following questions:

Panel Study

- In the cross-college reading and writing courses, what is the impact of the CTWC skills-building curriculum on students' levels of achievement over time?
- Does the early, deliberate teaching of the essential skills result in gains in learning the outcomes (the longitudinal research design)?

Course/Program Comparisons

- Across different student groups in different cross-college and program-specific courses, do diverse learning experiences in curricula related to CTWC result in different levels of skill achievement (the cross-sectional research design)?

Finally, a fourth objective of the study was to analyze the usability and scalability of the CTWC skills scorecard. Is the scorecard easy to use? Does it provide meaningful information to students? Is the CTWC skills scorecard measurement and assessment process scalable? To meet this end, the research team presents an analysis around the following questions:

Usability and Scalability

- Do faculty find the scorecard easy to use?
- Do faculty believe the scorecard will supply meaningful information to their students about their level(s) of achievement in the CTWC areas?
- Can the CTWC skills scorecard be implemented optimally within the institution and scaled successfully?

6. Literature Review

In 2014, Canada had the highest share among OECD countries of adults holding a tertiary education qualification (OECD, 2015). While the average level of attainment for OECD countries was 34%, 54% of Canadian adults between the ages of 25 and 64 held a tertiary qualification, and, of this group, a quarter had attained a diploma issued by either a community college or polytechnic institute as their highest qualification, a result higher than in any other OECD country.

Although this result speaks well of the PSE system in Canada, these attainment levels are often questioned in light of OECD rankings in literacy and numeracy for Canadian adults. Specifically, in 2012, the Programme for the International Assessment of Adult Competencies (PIAAC) indicated a ranking for Canada at the OECD average in literacy. However, a larger proportion of the Canadian population between the ages of 16 and 65 is at either the highest (14%) or lowest (17%) levels of literacy. In the assessment of numeracy, Canada ranks below the OECD average, and the proportion of Canadians between the ages of 16 and 65 at the lower levels of numeracy is greater than the OECD average.[2]

As a result of this growing divide between levels of postsecondary attainment and individual skill levels, PSE institutions have been tasked with redefining the ways in which they determine that students have the skills and knowledge necessary to be successful in, and contribute meaningfully to, their communities at local, provincial, national and international levels. This desire for greater transparency, coupled with a growing demand for increased accountability and quality, has given rise to the introduction of the OBE system. Within this system, educators can evaluate the quality of their programs by objectively assessing the attainment levels of their students; practitioners can adapt and refine their pedagogy to more effectively support student learning; and graduates can more confidently speak to the skills, knowledge and abilities they have acquired through their studies at the postsecondary level.

2 Additionally, the Programme for International Student Assessment (PISA), administered every three years by the OECD to measure the scholastic performance of 15-year-old secondary school students, indicated that the global ranking in mathematics for Canada dropped from seventh place in 2006 to tenth place in 2009 and, then, to thirteenth place in 2012. Rankings in reading remained stable between 2000 and 2012.

The Government of Ontario and Ontario colleges have championed the value of — and benefitted from — OBE for many years. The Ontario Qualifications Framework, introduced in 2009 by the then-named Ministry of Training, Colleges and Universities (MTCU), clearly identifies the competencies that a graduate must demonstrate at any given postsecondary credential level offered at one of the province's postsecondary institutions. These competencies include descriptors related to graduates' depth and breadth of knowledge, experience in research and scholarship, awareness of disciplinary or vocational concepts and methods, communications skills, ability to apply disciplinary or vocational knowledge, professional capacity and autonomy, and awareness of the limits of their knowledge (MAESD, 2009a).

Prior to the establishment of the Ontario Qualifications Framework, Ontario's publicly funded colleges were already focused on quality, accountability and accessibility through the College Standards and Accreditations Council (CSAC), which was established in 1993 and tasked with developing program standards for the entire college system. While CSAC ceased to exist in 1996, its work continues through various bodies within the Ontario Ministry of Advanced Education and Skills Development (MAESD), formerly MTCU. To date, over 200 program standards have been developed, approved and released, with a commitment to reviewing these standards on a cyclical basis in collaboration with the province's colleges (MAESD, 2016).

This move to establish program standards positioned Ontario at the forefront of OBE. However, Ontario did not remain focused solely on vocational learning outcomes. The ministry recognized that, to be successful, graduates needed to demonstrate the attainment of competencies beyond those of a specific vocational program as outlined in the Ontario Qualifications Framework. As such, for over a decade now, Ontario's colleges have required that all postsecondary college programs satisfy a general education requirement and build into every program curriculum a set of six skills categories: communication, numeracy, critical thinking and problem solving,[3] information management, interpersonal skills, and personal skills, known collectively as the EES (MAESD, 2003; revised in 2005).[4]

As noted, OBE is not new in Ontario and it has, in many ways, been supported over the previous two decades by the development of various policy documents that include explicit learning outcomes statements, and define skills that graduates must demonstrate in order to satisfy credential completion requirements (Liu, 2015). In addition, the general consensus is that there is widespread support at the institutional level to assist faculty and academic units in the development of learning outcomes-focused curricula and the evolution of teaching practices to better reflect and support OBE (Dawson et al, 2014, as cited in Liu, 2015). Furthermore, practitioners are increasingly encouraged to use a variety of assessment

3 The critical thinking and problem solving category specifies the following skills: analyzing, synthesizing, evaluating, decision making, and creative and innovative thinking. These are skills that graduates should demonstrate by applying "a systematic approach to solve problems" and using "a variety of thinking skills to anticipate and solve problems" (Appendix A) (MAESD, 2006).

4 According to Martini and Clare (2014, p. 7), these are the skills "that are in keeping with the attributes employers are increasingly demanding of university graduates" (Conference Board of Canada / CBC, 2013b).

strategies, from formative to summative, to assess discipline-specific skills and vocational knowledge (Lennon, 2010, p. 4).[5]

However, despite the focus on OBE as a method of assessing educational quality and value, it is still a relatively new model (Lennon, 2010, p.4). Since 1998, under the direction of MAESD, Colleges Ontario has administered the annual Graduate and Employer Key Performance Indicators (KPI) survey to evaluate the quality of the education provided by the Ontario colleges. The KPI survey provides information related to graduate employment, graduate satisfaction and employer satisfaction that could support the larger scale introduction of learning outcomes assessment tools, especially in the areas of basic and higher-order cognitive skills.[6] However, when assessing skills outcomes, particularly the higher-order cognitive skills such as CT and problem solving referred to by Weingarten (2014, Feb. 13), there is to date no available handbook or practitioner's guide.[7] In fact, not only is there no standard way of assessing these outcomes, according to Deller, Brumwell and MacFarlane (2015), there is a lack of agreement as to their very definition and/or the skills, traits or abilities that comprise them. For example, the CT model developed and refined by Robert Ennis (1985, as cited in Kaupp, Frank & Chen, 2014) focuses on induction, deduction and value judgements and is informed by inferences, observations, statements and assumptions. The Paul–Elder model developed by Paul (1993; Paul et al., 1993) and then refined by Paul and Elder (2001, as cited in Kaupp, Frank & Chen, 2014) introduces a series of intellectual standards: clarity, accuracy, relevance, logicalness, breadth, precision, significance, completeness, fairness and depth. These standards are then applied to different elements of reasoning such as, but not limited to: questions, points of view, information, inferences, concepts, implications and assumptions to arrive at a finite list of intellectual traits that include: intellectual

5 *Learning Outcomes Assessment: A Practitioner's Handbook* (Goff et al., 2015) (sponsored by HEQCO) provides a variety of assessment strategies to assist not only in the mapping of learning outcomes but also in their assessment.

6 In her study "A Fine Balance: Supporting Skills and Competency Development," Lennon (2010) concluded from her review of the 2008 KPI data that there was a high correlation between graduates' assessment of the education and training they had received and employers' assessment of these same skills. She further states that the data suggest that "employers are more satisfied with the skills of their newly graduated employees than are the graduates themselves — with the exception of the areas of CT, problem solving, research and analysis, organization, and planning" (2010, p. 19). In 2015, the provincial KPI graduate employment rate stood at 83.6%, the graduate satisfaction rate was lower at 80.3% and the employer satisfaction rate significantly higher at 91.4% (Colleges Ontario, 2016). The employer satisfaction rate in 2015 had also risen from where it was in 2014 (88.1%, Colleges Ontario, 2015). If employers indicate with such resounding clarity that they are and continue to be satisfied with the skills of graduates of Ontario's colleges, then there is no issue. However, when looking more deeply at the data, we cannot help but notice that the satisfaction ratings in specific skills areas differ from the overall satisfaction level. When graduates were asked about their satisfaction with their CT skills, 90.7% indicated that they were satisfied. By contrast, 88.5% of employers indicated they were satisfied with the CT skills of college graduates. At Humber College, the CT employer satisfaction rate decreased from 86.6% in 2014 to 86.2% in 2015, while the CT graduate satisfaction rate rose to 90.3% in 2015 from 89.9% in 2014.

7 In his discussion of learning outcomes, Weingarten (2014, Feb. 13) actually classifies learning outcomes into four different categories whose assessment is as varied as the categories themselves:

- Discipline-specific skills
- Basic cognitive skills: literacy and numeracy
- Higher-order cognitive skills: problem solving and critical thinking
- Transferable skills

humility, autonomy, integrity, courage, perseverance, empathy, confidence in reason and fair mindedness. For Pascarelli and Terenzini (2005), critical thinking as a skill is defined by one's ability to demonstrate some but not all of the following skills:

> identify central issues and assumptions in an argument; recognize important relationships; make correct references from the data; deduce conclusions from information or data provided; interpret whether conclusions are warranted based on given data; evaluate evidence of authority; make self-corrections; and solve problems (p. 156).

Bok (2006) on the other hand defines critical thinking as an "indispensable means of making effective use of information and knowledge" (p. 109) and, in so doing, expands the definition to include any process that allows for one to arrive at what he terms "carefully reasoned judgments."

This quick survey reveals the extent to which definitions of CT can vary. Understandably, this lack of consensus as to the elements that define critical thinking has had an impact on our ability to assess it as a skill. In fact, there is no single way in which CT is assessed. The CLA is a large-scale assessment tool which was designed to assess CT in a holistic manner as represented by analytic reasoning, problem solving, written mechanics and writing effectiveness (Kaupp, Frank & Chen, 2014), but, at present, the CLA lacks scalability due in part to issues of participant recruitment and administration (Lennon, 2014).[8]

These are just a few of the challenges that Ontario's postsecondary institutions, specifically Ontario's colleges, face as they continue to evolve processes and practices to address issues of accountability, quality and value. In no area are these issues more contentious than when assessing CT, one of the EES that graduates must demonstrate in order to meet the requirements of their postsecondary credential, and one of the most highly valued skills among employers of Ontario's college graduates.

7. Methodology

7.1 The CT Construct

The Foundation for Critical Thinking defines the skill of CT as an "intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning or communication, as a guide to belief and action" (Scriven & Paul, 1987). When considered this way, CT has two discrete components: the skill to analyze or evaluate information and the skill to use analytic or evaluative skills to guide behaviour.

8 Lennon (2014, p. 7) further states that the development of standardized learning outcomes assessment tools raises an additional concern that the results may be used as ranking tools instead of contributing to accountability, quality and mobility.

When considered in an academic setting, CT can be better understood to include "the ability to analyze a piece to assess its quality, often by comparing it to other instances or by breaking it down according to some measure of quality" (Foundation for Critical Thinking, 2015, as cited in Joordens, Paré & Collimore, 2014, p.5). However, as a skill, it must be practised repeatedly and "preferably within a context that provides structure and guidance" (Joordens, Paré & Collimore, 2014, p. 6). For this reason, CT is difficult to both teach and assess.

As detailed earlier and evidenced by its inclusion as one of the EES required of all graduates of a college program (MAESD, 2003; revised 2005), CT is considered an important outcome of a postsecondary education. However, according to Arum and Roksa (2010), not only are we not assessing this skill set, students are not making the necessary gains in CT. Indeed, few PSE institutions require students to take mandatory courses focused specifically on CT. In fact, most students are expected to develop CT skills through taking a variety of courses that inherently foster these skills and completing assignments where they are evaluated on how well they demonstrate CT skills. Similar to most other institutions, Humber College has very few courses where CT is taught explicitly. Instead, CT is incorporated into its postsecondary courses as one of the EES that students are required to demonstrate upon graduation. As such, CT may variously be taught, reinforced, and/or assessed as an integral component of a program of study.

At Humber College, one course where CT is explicitly taught and reinforced as an essential skill is the first semester reading and writing course, delivered by the Department of English to all certificate and diploma students. The WRIT 100 course (formerly COMM 200), focuses on developing students' ability to read carefully, write effectively and think critically.[9]

Another context where CT is emphasized, if not explicitly taught, is in certain vocational programs, such as Practical Nursing and Police Foundations, where industry expectations demand that graduates be equipped to work with vulnerable populations or hold positions of authority.

Last, in addition to the skills-based approach to teaching CT, the GAS programs at Humber all include a dedicated course titled Critical Thinking as part of their core curriculum, where CT is taught as content knowledge.

For the purposes of this study, we have assessed the CT and WC skills in each of the above contexts to determine which groups of students demonstrate the greatest abilities: those who explicitly learn CT as an essential skill; those for whom the skill is more implicitly embedded and reinforced in program content as a requirement of their chosen vocational field; and, those who take a content course on CT.

9 Through a metacognition-focused pedagogical approach, critical reading, argument analysis and critical response are used as the context within which to teach students the process of critical thinking; these skills then form the learning outcomes for the course and are assessed accordingly (Appendix E: Course Curriculum).

7.2 Instrument Development and Description, Testing and Redesign

During the first phase of the study in the academic year 2013–2014, steering and pilot committees at Humber College developed the CTWC skills scorecard, and trained faculty to evaluate students' work in two cross-college courses (COMM 200: College Reading and Writing Skills and COMM 300: Workplace Writing Skills), and in the Police Foundations, Business Administration, and Business Management diploma and advanced diploma programs, specifically in the courses PFP 301: Criminal and Civil Law, PFP 403: Police Oversight, BMGT 300: Human Resource Administration, BMGT 500: Strategic Management, BFIN 420: Business Finance, BFIN 500: Corporate Finance, and BACC 300: Financial Analysis and Accounting.[10] Over 1,000 students participated in the pilot study.[11]

Humber's original CTWC skills scorecard (Appendix B), was constructed to assess the cognitive skills from the American Philosophical Association's (APA) consensus definition of CT, derived from the rigorous Delphi research method (APA, 1990; Facione, 1990). Referred to as the Delphi definition, it stresses CT as: "[the] purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgement is based" (Facione, 1990, p. 2). As such, the initial scorecard consisted of six items in critical thinking (interpretation, analysis, evaluation, inference, explanation and self-regulation) and three in written communication (structure, style and mechanics).[12] Within each item there was a seven-point level of achievement scale with the following categories: 0 (no evidence), 1 (poor), 2 (fair), 3 (good), 4 (very good), 5 (excellent), 6 (mastered). Students' CTWC skill attainment was represented as numeric scores out of six on each of the nine items of the rubric.

Faculty tested the CTWC skills scorecard by applying it to assignments in the aforementioned courses and student groups (typically one assignment in one course per semester). As depicted in Table 1, students demonstrated, on average, good levels of total CT and WC skill achievement in more advanced reading and

10 The project was initially launched at Humber College's Lakeshore Campus using an open call for participation. After the initial expressions of interest were received, the steering committee then worked with each academic school to identify the participating programs and courses that would provide a range of learning contexts within which CTWC could be measured. For example, the School of Social and Community Services identified the Police Foundations program because of the critical thinking imperative that exists within that context. In contrast, the School of Liberal Arts and Sciences writing courses were selected because of the opportunity afforded to examine a context where CT skills are explicitly taught and evaluated.

11 In fall 2013, 434 students were included in the pilot phase. In winter 2014, 813 students participated. However, final sample sizes were lower for reasons such as students dropping the course or not completing the course evaluation components. For example, the valid sample size was 363 students for Fall 2013 and 636 students for Winter 2014. For details of the pilot study results, consult Humber College (2014) and Humber College (2013–2014).

12 Grof et al. (2015, p. 31) state that it is not uncommon for the assessment of CT to occur concurrently with the assessment of writing in assignments that are of an argumentative, persuasive and evaluative nature. This congruency is reflected in the first scorecard designed and piloted in this study.

writing courses (COMM 300) and business courses (BMGT 300, BMGT 500, BFIN 420). In all other courses, students typically achieved fair level of total CT and WC skills scores.[13]

Table 1: Differences in Total Scores on CTWC Skills by Course (2013–2014)

Course	Critical Thinking		Written Communication		
	M	SD	M	SD	N
Fall 2013					
COMM 200	2.41	0.89	1.95	0.81	134
PFP 301	2.78	0.98	2.88	0.95	171
BMGT 300	2.49	1.09	2.53	0.70	33
BMGT 500	3.30	0.45	3.30	1.01	23
BFIN 500	2.30	0.76	2.80	1.21	10
Winter 2014					
COMM 200	2.87	0.90	2.73	0.88	154
COMM 300	3.15	1.03	3.26	1.00	141
PFP 403	2.65	0.74	2.63	0.84	205
BACC 300	2.66	0.87	2.93	0.80	40
BMGT 300	3.47	0.93	3.30	1.10	24
BMGT 500	3.31	0.44	3.25	0.59	19
BFIN 420	3.65	0.74	3.80	0.73	25
BFIN 500	2.39	1.23	2.89	1.09	21

Source: Humber College (2013–2014) Note: M = Mean; SD = Standard Deviation; N = Sample Size
Total CT and WC skills scores range from 0 (no evidence), 1 (poor), 2 (fair), 3 (good), 4 (very good), 5 (excellent) to 6 (mastered).

The pilot phase also included an assessment of the reliability and validity of the original scorecard. Data analysis revealed the initial scorecard had high reliability (internal consistency) and moderate validity (concurrent validity) (Humber College, 2013–2014, pp. 18, 20).

Further, in a separate questionnaire, faculty were asked whether they found the scorecard easy to use and whether they believed the scorecard would supply meaningful information to their students (Appendix D). Faculty were also asked whether they would be willing to join a focus group to discuss the scorecard.

13 Notably, in the Winter 2014 semester, students in both reading and writing courses (COMM 200 and 300) had higher average total CT and WC skills scores than students in the more senior police foundations course (PFP 403) and some business courses (BACC 300, BFIN 500).

Overall, the data from the pilot study revealed some interesting initial distinctions between the courses and student groups, which informed the development of our hypotheses in the second phase of the study. The steering and pilot committees also learned some valuable methodological lessons from the analysis of both the student and faculty data sets that led them to contemplate scorecard redesign.

Accordingly, after the pilot study, the committees consulted with faculty across campuses and programs to solicit feedback about their initial assessment experiences, including suggestions to redesign and improve the instrument. Focus groups were also held to discuss specific ideas for the revision of the scorecard. Some faculty members found the scorecard quick and easy to use. Others found it time consuming as they were not as familiar with the meaning of each field. However, the most frequently reported complaint was that a scorecard with separate fields for CT and WC was unnecessarily granular because the CT fields, in fact, had the expression of the WC fields embedded in the demonstration of CT (M. Gamble, Personal Communication, November 21, 2013).

Not only did faculty indicate that the tool was too complex and included too many criteria and fields (R. Hussein, Personal Communication, October 1, 2013), there was consensus that the formal assessment of both CT and WC skills on a single scorecard seemed to lack congruency (M. Gamble, Personal Communication, November 21, 2013). Indeed, data analysis indicated that students' CT scores did not substantially differ from their scores on WC (Humber 2013–2014). Over the course of the pilot year, there was also evidence of confusion surrounding several rubric descriptors; for example, there was divergent understanding of the differences between evaluation and inference (E. Harvey, Personal Communication, July 3, 2014). As a result, the steering committee made refinements to the items of the skills scorecard to ensure that it was simpler and easier to use, universal for faculty, and feasible for data collection and reporting.[14]

Specifically, the first two CT descriptors were categorized under a dimension newly named "comprehension" and the last four were categorized under the dimension "integration of writer's ideas." The three WC descriptors were then removed entirely and writing was embedded in the CT descriptors. Last, two levels of achievement were removed from the skills scorecard: no evidence (0) and mastered (6). The revision process resulted in a simplified scorecard decreasing from nine items to six with increased agreement among the committee and faculty that the remaining six items belonged best in the scale. All in all, the research team's changes made the tool easier for faculty members to use and better reflect the fact that reading, CT and writing are inextricably linked, particularly since reading is the "input" and writing is the "output" (E. Harvey, Personal

14 The challenge of distinguishing CT from WC is not new and is further enhanced by the fact that several assessments of critical thinking like the Collegiate Learning Assessment (CLA), actually measure problem solving and communication skills as a part of assessing CT, However, in this study, the revisions to the scorecard do not reflect this challenge, but rather a need on the part of the steering committee to make the tool more user-friendly, especially for faculty who are not accustomed to assessing written language skills. For a full discussion of the content, validity and reliability of the CLA, please see Benjamin (2013).

Communication, October 26, 2014). In many aspects, the revised descriptors also better reflected the Delphi definition and the relevant CTWC constructs.

7.3 The Revised Instrument

The revised CTWC skills scorecard (Appendix C) contains six items under two headings: comprehension and integration of writer's ideas:

Comprehension

- Interpretation
- Analysis

Integration of Writer's Ideas:

- Evaluation
- Inference
- Explanation
- Self-regulation

Within each item, there is a five-point level of achievement scale with the following categories: 1 (poor), 2 (fair), 3 (good), 4 (very good) and 5 (excellent). Students' CTWC skills attainment can be represented as numeric scores out of five on each of the six items of the rubric. Sums of total scores can also be computed and categories for sums of total scores can then be collapsed.

Since the CTWC skills scorecard outlines five levels of achievement, the assumptions behind the study, compatible with the Ontario Qualifications Framework detailed earlier, are that three-year advanced diploma graduates should have reached achievement level 5 (excellent); two-year diploma graduates should cultivate skills commensurate to level 4 (very good); and one-year certificate graduates should have acquired skills at level 3 (good).[15] Accordingly, expectations for progress over the full range of students' levels of learning in their programs until exit are the following:

15 Students who do not meet the standard for WRIT 100 presumably have poor (level 1) CTWC skills ratings. It should be noted that there is a college entrance placement testing of basic academic literacy for all students, and those who do not meet the departmental standard for entry into first-semester college writing go into an alternative course WRIT 050: Introduction to College Writing Skills (formerly COMM 100: Introduction to College Reading and Writing Skills). To complete WRIT 050 successfully and move into WRIT 100, students must produce writing that meets the minimum departmental standards for entrance. Once in WRIT 100, they engage with the critical analysis curriculum presumably leading, at a minimum, to fair (level 2) CTWC skills assessments by the end of the course. Here, the rationale for assigning the poor (level 1) rating to WRIT 050 is based on the assumption that students with below-college literacy do not have the WC skills to adequately communicate CT processes and produce work at a higher level on the CTWC skills assessment. Thus, WRIT 050 is about levelling the playing field so that students can engage alongside WRIT 100 direct entrants to achieve fair or higher CTWC skills upon exit. Accordingly, this is a remedial strategy to address the needs of students whose CTWC skills are poor.

- Semester 1: scores of 2 (fair) or above
- Semester 2: scores of 3 (good) or above (certificate graduates)
- Semesters 3 and 4: scores of 4 (very good) or above (diploma graduates)
- Semesters 5 and 6: scores of 5 (excellent) (advanced diploma graduates)[16]

Thus, the revised scorecard is not only more straightforward and easy to use with fewer items and a simplified scale, it is also better designed to assess the learned CTWC skills of college-level students, including their ability to comprehend and integrate writer's ideas into their work as they move through their certificate, diploma and/or advanced diploma programs. This said, faculty have not, as yet, decided upon exact expectations for student skills scores for lower-level and higher-level courses and semesters. As such, the approach to assessing learning outcome expectations above should be considered an ideal-type classification that can enable longitudinal and comparative analysis of student work and skills scored using the instrument at or above each of the milestone levels of achievement (Siefert, 2011).[17]

With a new CTWC skills scorecard, we needed to reassess its effectiveness by examining its validity and reliability (Finley, 2011; Pusecker, 2011). The key types of validity discussed and/or analyzed are content and face validity, criterion validity (concurrent) and construct validity (convergent). The main forms of reliability assessed are inter-rater reliability and internal consistency. Using the same survey from the pilot phase, faculty were asked about their perceptions regarding the usability and meaningfulness of the revised scorecard (Appendix D). All findings are presented in the results section.

7.4 Study Design: Course Typology

During the second phase of the project, which took place over the academic years 2014–2015 and 2015–2016, faculty used the revised scorecard to assess the CTWC skills of students in a variety of courses where these skills are either explicitly taught, more implicitly embedded, or focused on as discipline content. In order to better understand the study design, it is important to understand the instructional context and course typology.

Like so many of their counterparts at other Ontario colleges, first semester students at Humber College are required to take a cross-college reading and writing course focused on enabling them to practice their reading and writing skills while also earning credits towards their postsecondary credential (Fisher & Hoth, 2010).

16 Similarly, in the AACU VALUE rubrics in CT and WC there are four levels of achievement: not met, benchmark (level 1), two levels of milestones (levels 2 and 3), and capstone (level 4). The assumption is that level 4 of the rubric is the characteristic that schools want graduates to demonstrate in their assessed work products by the completion of their programs. Thus, for the 100 level, the expectation is for students to score at or above level 1 (the benchmark), for the 200 level at or above level 2 (the first milestone), for the 300 level at or above level 3 (the second milestone) and for the 400 level at level 4 (the capstone) (Greenhoot & Bernstein, 2011).

17 If desired in the future, a level 6 (mastered) can be readily incorporated into this extant framework to evaluate the acquired CTWC skills of four-year baccalaureate's/bachelor's degree and honours degree students upon graduation.

As per the course description, this foundational course, COMM 200: College Reading and Writing Skills, emphasizes the essential elements of the reading and writing process (Appendix E).[18] Students practice reading and writing skills required to succeed in both their college programs and in professional communication. To reach these goals, the course covers the following: critical reading and thinking; paraphrasing and summarizing; the writing process; critical response; persuasive writing; revision; grammar and mechanics; and documentation. In the fall of 2014, with the introduction of a new student information system, COMM 200 was recoded as WRIT 100. Some aspects of the course continued unchanged, while other important curriculum changes were made. For example, the course evaluation methods now require students to practice and produce critical analysis response essays instead of writing assignments demonstrating their understanding of the different rhetorical modes, which had been the case in COMM 200. In addition to the production of critical analysis response essays, students also complete at least one revision requiring that they engage more meaningfully with metacognitive techniques. With a renewed focus on analysis, evaluation, synthesis, explanation and metacognition, WRIT 100 requires that students demonstrate the cognitive skills defined through the Delphi research method as being at the core of CT (APA, 1990; Facione, 1990).

In almost all diploma programs at Humber College, students are not only required to take a first semester writing course that introduces them to postsecondary level reading and writing expectations; they are also required to take a second semester course that focuses on writing in a professional workplace context. In 2013–2014, this course was called COMM 300: Workplace Writing Skills and it introduced students to the strategies of effective written workplace communication. The course was designed to build on and reinforce the writing skills developed in COMM 200, and requires students to apply these skills to vocationally relevant assignments where they need to demonstrate the ability to select and organize pertinent information according to purpose and audience, and present their ideas clearly, precisely and effectively in various written formats. In the Fall 2014 semester, like its predecessor COMM 200, COMM 300 was given a new course code: WRIT 200. Even though the course title remained unchanged, the course itself underwent a significant revision to better align it with the new focus of WRIT 100 and to better reflect the changing priorities of the Department of English, namely the renewed focus on interpretation, analysis, evaluation, inference and explanation.

As listed in the course outline, WRIT 200 “builds on and reinforces the critical reading, thinking, and writing skills learned in WRIT 100. Students will broaden their concept of audience and refine their ability to

18 In 2013–2014, when the first phase of this study was completed, most students were required to take COMM 200. They were permitted to register upon completion of a standardized, post-admission placement test that required them to produce an essay graded by Department of English faculty. In some instances, students did not demonstrate the minimum level of proficiency to register for COMM 200. In those cases, they were required to take a non-credit, remedial course, COMM 100: Introduction to College Reading and Writing Skills (now WRIT 050: Introduction to College Writing Skills), before proceeding to the required credit course in their program of study. In other instances, this same placement test identified students who had already mastered learning outcomes of COMM 200. On these rare occasions, students were exempted from taking the course. This test also allowed students to be placed in a parallel stream of English courses with a focus on supporting the needs of students whose first language is not English. This post-admission placement testing is still in effect; however, students now complete an online essay through WritePlacer, supported by The College Board.

synthesize information from various sources" (Humber College: WRIT 200 Course Outline, 2014) (Appendix E). To help students meet the challenges and expectations of communicating in a changing workplace, they are required to produce written pieces using different workplace formats; analyze audience and purpose when producing written work; research, evaluate and synthesize the ideas of others; and problem solve (Humber College: WRIT 200 Course Outline, 2014).

While CT skills are not taught as overtly in WRIT 200 as in WRIT 100, the WRIT 200 curriculum deliberately builds on the skills developed in WRIT 100. This would imply that measuring CT gains in students who completed WRIT 100 and then progressed through WRIT 200 should reveal more significant gains than in students assessed only on the completion of WRIT 100, which is the focus of the panel study in this research project.

In addition to assessing the CTWC skills of students in courses where CT is explicitly taught as an essential skill, this study also focuses on assessing the CTWC skills of students in a program of study where the industry expectation is that graduates will be required to actively use the skill. In this case, the program selected is the two-year PFP at Humber College where students take a variety of courses requiring them to apply CT as they examine course content and learn new skills. For the purposes of this study, the two courses in question are PFP 200: Criminal and Civil Law, and PFP 211: Interviewing and Investigations, both taught in the third semester of the program. In each of these courses, second year students need to demonstrate their knowledge and understanding of the criminal law system with a specific focus on the role of criminal investigation and enforcement as a patrol officer, and with an emphasis on their legal responsibilities as outlined in the Charter of Rights and Freedoms (Appendix E). In addition, PFP 211 students focus on developing their interviewing and investigation skills both in the context of attaining the skills necessary to retrieve information from witnesses, victims and suspects, and acquiring effective techniques for detecting deception (Appendix E). Given this focus on both the theoretical and practical elements of policing, students are required to demonstrate a variety of learning outcomes that require that they summarize, explain, evaluate and interpret (Humber College: PFP 200 Course Outline, 2014). They must also act in a manner that is consistent with relevant law and legislation as well as professional, organizational and ethical standards, while persuading others using a variety of communication skills in order to extract pertinent and relevant information (Humber College: PFP 211 Course Outline, 2014).

The last group of students whose CTWC skills were assessed were students required to take a discrete course in CT, GCRT 100: Critical Thinking, a core course in the GAS College, University, and Health and Science certificate and diploma programs. Unlike students exposed to CT as a skill, the GAS students are expected to examine CT from both theoretical and applied perspectives. In studying CT theory, students are required to consider the structure of arguments and examine the common fallacies of reason, the different forms of constructing an argument, and the rhetorical devices used to mask a bad argument (Humber College: GCRT 100 Course Outline, 2015) (Appendix E). When applying CT skills, students in GCRT 100 are required to:

- Describe the structure of arguments
- Identify premises and conclusions
- Analyze the relationships between premises and conclusions
- Distinguish between inductive and deductive arguments and patterns
- Evaluate arguments for validity, soundness, strength and cogency
- Judge the probability of claims of cause and effect
- Recognize fallacies of reasoning and construct arguments that satisfy the criteria of good argumentation

(Humber College: GCRT 100 Course Outline, 2015).

In each of these courses, one assignment was selected in which students' CTWC skills were assessed. In one instance, it was an in-course written assignment that requires students to read, interpret and analyze a written text before producing their own piece of writing: WRIT 200 (Appendix F). In another case, the written assignment was a research essay that requires students to research, evaluate, synthesize and accurately cite scholarly and popular sources to support their own persuasive writing: PFP 211 (Appendix F). In another example, students are required to analyze and interpret case law in order to produce a concise summary of a case law decision: PFP 200 (Appendix F). And in some cases, the assessment of students' CTWC skills occurred at the end of the course in the form of a final summative exam. In WRIT 100, this required students to read, analyze, interpret and evaluate an article in order to write a critical response essay (Appendix F). In GCRT 100, students were required to respond to a series of questions that required them to recall course content while also producing their own cogent arguments that satisfied the criteria of good argumentation (Appendix F). In all instances, students were required to demonstrate self-regulation on each assignment or exam, either by following explicit instructions to revise and edit their work before submitting for grading — as is the case with the WRIT 100 final exam — or by applying revising and editing steps learned in mandatory English and communications courses prior to entering their third semester, as is the case with PFP 200 and 211.

7.5 Study Design: Hypotheses and Variables

By administering the revised CTWC scorecard in cross-college reading and writing courses where the skills are explicitly taught (WRIT 100 and 200), in courses related to CTWC in the PFP diploma where skills are embedded (PFP 200 and 211) and in the GAS certificate program, where CT is the focus of the course content (GCRT 100), the research team was able to assess students' CTWC skills, as demonstrated in their written work, in several learning contexts. Selecting the cross-college reading and writing courses allows assessment of the early, explicit CTWC skills-building curriculum on changes in the learning outcomes of a small panel group of students over time. Adding the PFP and GAS courses and student groups enables the comparison of different approaches to teaching the skills in the diverse curricula, and an analysis of their effects on levels of achievement. Here, we tentatively hypothesize that the early, explicit CTWC skills-building student experience in the two cross-college reading and writing courses will result in gains in learning the skills over the two semesters, as well as in comparatively overall higher levels of achievement

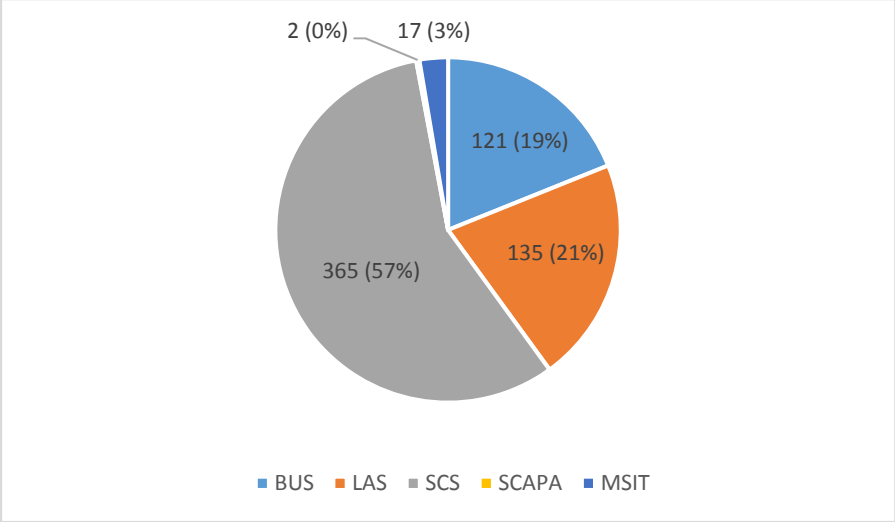
than in the more implicit approach to teaching the skills in the PFP program-specific courses and the GAS content-heavy CT course. In the first case, we examine the relationship between the explicit skills-building curriculum and students' CTWC skills scores over time. In the second case, we explore the relationship between the different curricula and the associated diverse student learning experiences in the courses, and different levels of achievement on students' CTWC skills scores. Here, we certainly recognize that the threats to internal validity in the longitudinal and cross-sectional research designs of the study are much higher than in true experiments and thus we should be cautious in arriving at any conclusions about causality.

7.6 Study Context: Participants and Sample-selection Procedures

A total of 650 students took part in the second phase of the study primarily from three of the college's eight schools: Business (BUS) (19%), Liberal Arts and Sciences (LAS) (21%), and Social and Community Services (SCS) (57%) (Figure 1). LAS houses the GAS College and University Transfer certificate program and SCS runs the diploma PFP. The remaining 3% of students were from the schools of Creative and Performing Arts (SCAPA), and Media Studies and Information Technology (MSIT). In total, students participating in the study were from 16 certificate, diploma and advanced diploma programs.[19] The skills scorecard was used by 46 faculty members to assess 728 student work products (417 final exams and assignments in the reading and writing and CT courses and 311 assignments in the PFP courses).

19 The programs involved were: Accounting (Diploma), Advertising and Marketing Communications (Diploma), 3D Animation (Advanced Diploma), Business Administration: Accounting (Advanced Diploma), Business Administration Co-op (Advanced Diploma), Business Administration (Advanced Diploma), Business Management: Financial Services (Diploma), Business Management (Diploma), Business Marketing (Diploma), Child and Youth Care (Advanced Diploma), Community and Justice Services (Diploma), Graphic Design (Advanced Diploma), Jazz Performance: Introduction to Commercial Jazz (Ontario College Certificate), Police Foundations (Diploma), Theatre Arts Performance (Diploma), and GAS: College Transfer (Ontario College Certificate) or University Transfer (Ontario College Certificate or Diploma).

Figure 1: Participating Students by School (2014–2016)



Most students and all faculty were chosen for the study on the basis of convenience. For example, students in the selected course sections were simply required to complete the noted assignments that were then assessed using the CTWC skills scorecard (low external validity).[20] Smaller subsets of students required for aspects of the validity and reliability analyses were further selected randomly from the student groups. Importantly, in the case of the non-random sample-selection procedures, we are careful not to generalize the results of the study beyond the student and faculty groups we analyze.

The Cross-sectional Group

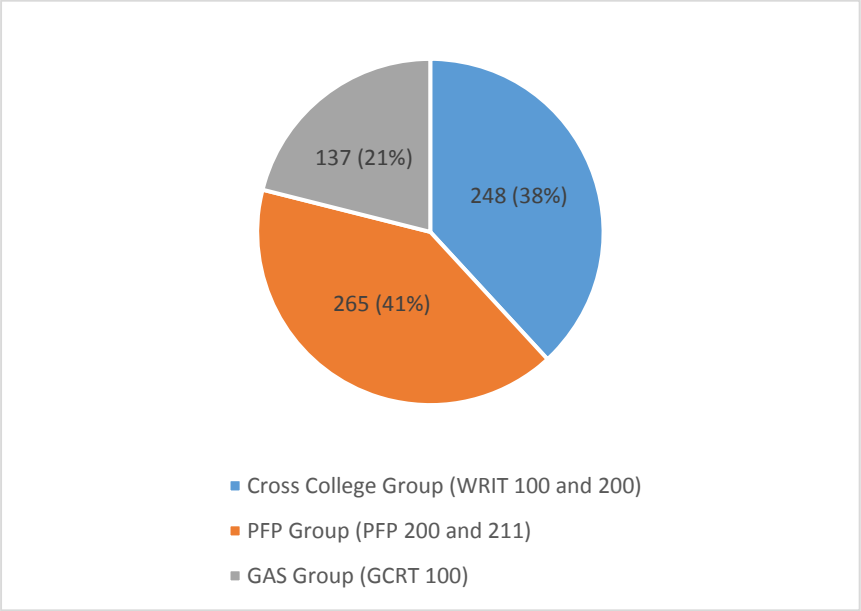
Students in every certificate and diploma program in arts, business and the social sciences at the college take the same reading and writing courses (WRIT 100[21] and 200), typically in their first and second semesters, where, effective Fall 2014, the essential CTWC skills are explicitly taught so students can actively learn and apply them. Of the 650 participating students, 38% fell into this cross-college group and were part of the CT as an explicit skills-building learning experience (Figure 2). Students in the diploma PFP take the

20 A full research risk assessment was conducted by Humber College's Office of Applied Research and Innovation and, as a result, the study was exempted from the requirements of formal research ethics review under the provision of Article 2.5 of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (2014) that governs quality assurance and quality improvement studies, program evaluation activities and performance reviews.

21 Of the 157 students who completed WRIT 100, 150 students completed the course in their first semester (the other seven students took the course in subsequent semesters of their programs). Further, 143 students out of the total 157 were taking WRIT 100 as a first attempt. Eight students were repeating the course, and two students were taking WRIT 100 as a first attempt after having completed the remedial writing skills course (WRIT 050). Data regarding attempts were missing for the remaining four students.

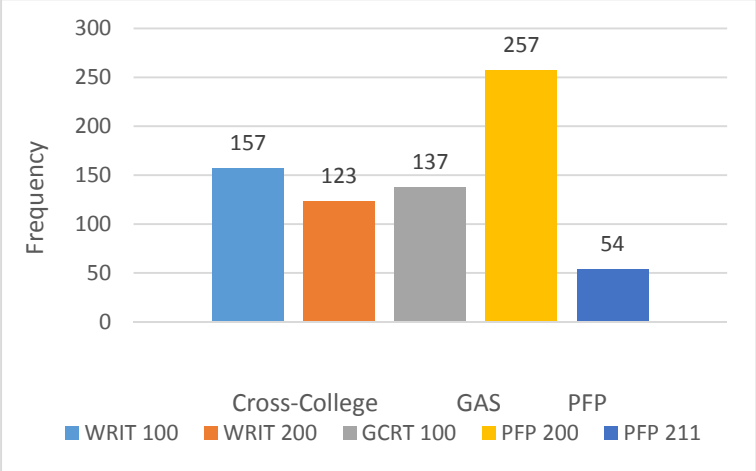
criminal and civil law (PFP 200) and interviewing and investigation courses in their third semester (PFP 211); in these courses, CT concepts are embedded but the essential components of the skill are not explicitly or deliberately taught. Forty-one percent of student participants were in this group, learning CT more implicitly or indirectly in a program-specific context. Twenty-one percent of students were taking a CT course that emphasizes learning and applying conceptual and theoretical knowledge related to CT rather than the explicit building of the components of the skill. This course normally occurs in the second semester of the GAS certificate programs.

Figure 2: Participating Student Groups by Course (2014–2016)



Importantly, 32 students in the cross-college WRIT courses had their student work assessed twice (final exams in WRIT 100 and assignments in WRIT 200) (5% of the total 650 student sample). Similarly, 46 students in PFP took both the 200 and 211 courses and thus had their assignments assessed twice (7% of the student sample). The 728 student work products assessed by course for the second phase of the study are presented in Figure 3.

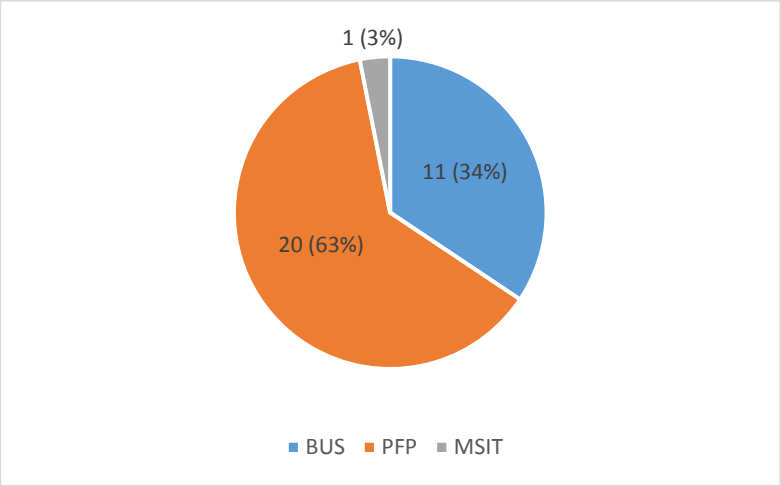
Figure 3: Student Work Products Assessed by Course (2014–2016)



The Panel Group

The panel group consisted of 32 students who took the cross-college reading and writing courses back-to-back in the fall and winter semesters of the 2014–2015 academic year. As such, their work products were assessed in both courses. In spite of the relatively small size of this panel group, three schools at Humber were represented as were both diploma and advanced diploma students. Thirty-four percent of students in the panel group were enrolled in diploma and advanced diploma business programs; 63% of students were registrants of the PFP diploma in SCS; and one student was from the diploma program Advertising and Marketing Communications in MSIT (Figure 4).

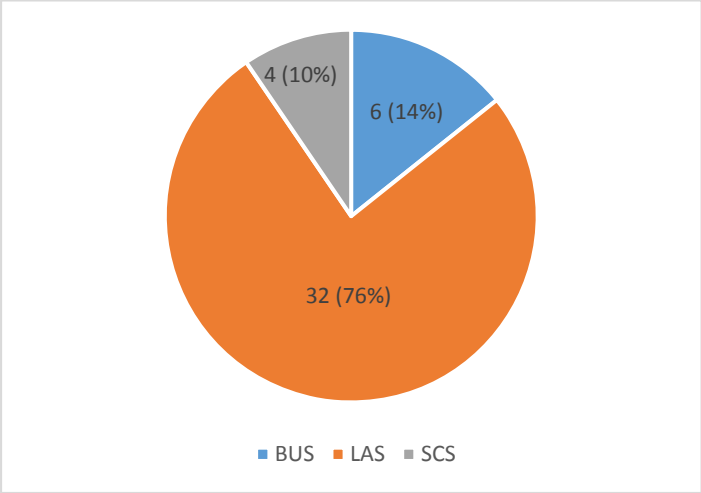
Figure 4: Participating Panel Students by School (2014–2015)



The Faculty

A total of 46 faculty from the three main schools participated in the two phases of the study. The faculty participant breakdown was 76% from LAS, 14% from BUS and 10% from SCS (Figure 5).

Figure 5: Participating Faculty by School (2013–2016)



7.7 The Administration of the Skills Scorecard

Once the courses in which the scorecard would be administered were identified, faculty were invited to participate in the study and use the scorecard on an existing piece of in-course written work that required students to demonstrate the cognitive skills associated with CT: interpretation, analysis, evaluation, inference, explanation and self-regulation (APA, 1990; Facione, 1990). In addition, faculty were directed to ensure that the following conditions were met when selecting an assignment for student assessment:

- Assignment requires pre-reading
- Assignment written in class
- Assignment carries a grade value of at least 15%
- Assignment is given in the last third of the course
- Assignment is individually authored (HEQCO-LOAC Steering and Pilot Groups, 2014) (Appendices F and G).[22]

22 In the summer of 2014, these criteria were revised slightly, namely there was no longer a requirement for pre-reading and the assignment only had to be worth at least 10% of the final grade. In addition, faculty were directed to ensure that the assignment required reading, writing and CT (E. Harvey, Personal Communication, August 24, 2014).

As detailed earlier, in some instances, faculty selected a formative assessment such as a written assignment, essay or case brief as was the case in WRIT 200, PFP 200 and PFP 211. In the case of WRIT 100 and GCRT 100, the skills scorecard was administered to the final exam (Appendix F).[23] In all cases, the skills scorecard was administered post midterm ensuring that students would have had sufficient time to explore new content and practice the skills that were to be assessed. Additionally, the scorecard was administered only once and after the actual grade for the assignment was determined. This ensured that there was no risk of the student's grade being influenced by the CTWC score.

7.8 The Training and Scoring Protocol

With a project of this magnitude that spanned several programs, schools and campuses, the only way to ensure consistency of use was to provide formal orientation and training for faculty who elected to participate. These orientation sessions were facilitated by members of the steering committee and held at the start of each semester to ensure that faculty were able to identify appropriate assignments with a clear understanding of the project and skills scorecard. For those unable to attend, electronic copies of the faculty information package were distributed by steering committee members with an open invitation to continuously share feedback both formally and informally, as individual faculty administered the tool and returned their scorecards and faculty surveys to the research office for tabulation and analysis by the project team (Appendix G).

8. Results: Data Presentation and Interpretation[24]

8.1 Validity and Reliability Results

Validity

Does the CTWC skills scorecard measure what it is intended to measure? Do the individual indicators of the skills scorecard devised to measure CTWC really measure those concepts (construct validity)? Do experts agree that the individual indicators and overall measure reflect the concepts? Do faculty in the field agree (content and face validity)?

23 In the case of GCRT 100, the faculty were not involved in the discussions regarding the development of the original scorecard or its revised version that was used in 2014–2016. Given their absence from these foundational conversations, they were not asked to administer the scorecard but instead a third party assessor who had used the scorecard in his or her own WRIT courses was engaged.

24 All data in the second phase of the study were anonymized and then analyzed using IBM SPSS Statistics 23. We primarily relied on correlational statistical procedures to establish the validity and reliability of the revised CTWC skills scorecard: Pearson's r and Spearman's rho correlation coefficients (convergent validity), the two-way, mixed, consistency intraclass correlation coefficient (ICC) (inter-rater reliability (IRR)), and Cronbach's alpha and item-total correlations (internal consistency). A more detailed presentation of the data analysis and interpretation is available in Appendix H.

The CTWC skills scorecard development process at Humber, which involved the steering and pilot committees and faculty close to student learning and outcomes assessment across campuses, schools and programs, established a credible extent of face validity (Finley, 2011, p.1). In particular, after the first pilot phase, campus consultations resulted in a revised, simplified scorecard with increased consensus that the remaining six items belonged best in the scale and the overall measure better reflected the Delphi definition and the CTWC constructs concerned. Further, the foundation of this study and the scorecard itself is based on the Delphi method that APA experts used to arrive at the consensus definition of the CTWC constructs, which provides the skills scorecard with a substantial degree of content validity (Facione, 2013; Saxton et al., 2012). Thus, we suggest the development, testing and redesign processes detailed in earlier sections provided the CTWC skills scorecard with substantial amounts of content and face validity.

However, the validity of the skills scorecard is not solely determined by whether or not the instrument appropriately captures the full range and richness of the concepts involved. In fact, its validity stems equally from its ability to accurately measure actual skills achievement. As a result, we asked a series of additional questions to continue to assess its validity.

How well do students' CTWC skills scores compare with other established measures (criterion validity)? Do students' CTWC skills scores correlate with their scores on an established CT skills measure such as the TER (concurrent validity)?[25]

During the initial phase of the study, the concurrent validity of the original scorecard was investigated. A randomly selected subset of 54 students enrolled in college writing (COMM 200, 300), business (BMGT 300, 500) and police foundations (PFP 301) courses were assessed. Both the original CTWC scorecard and the TER, a member of the California Critical Thinking Skills Test (CCTST) family of CT skills tests were applied to their in-class assignments. A comparison of results revealed a significant, positive and moderate correlation between students' total scores on the CT component of the original skills assessment tool and their corresponding CT scores on the TER (Humber College, 2013–2014, p. 20). As such, this initial concurrent validity investigation provided preliminary evidence for criterion validity. However, since validity should not be established in studies with a singular measure, we felt it would be wise to perform additional convergent assessments of an alternative form of validity — construct validity — in the second phase.

Do students' CTWC skills scores correlate with other theoretically similar constructs such as grades on assessed work and final course grades (convergent validity)?

25 It is important to note that the study did not control for any variables as the primary focus of the project was to determine the usability and scalability of the CTWC scorecard versus examining and/or analyzing the many variables that could impact student achievement in CTWC.

In the second phase of the study, students' total CTWC skills scorecard results in the reading and writing (WRIT 100) and CT (GCRT 100) courses were also compared using the appropriate correlation coefficients with final grades on the associated exams.[26] We reasoned that large correlations between scores on the revised CTWC skills scorecard and scores on the course exams that are explicitly designed to assess the same learning outcomes would provide evidence of convergent validity. Data analysis found that the total CTWC skills scores were positively and largely correlated with final exam grades in both WRIT 100 and GCRT 100 providing evidence of the convergent validity, and hence construct validity, of the scorecard.

When students' total scores on the CTWC skills scorecard were further correlated with their final course grades, positive and large to moderate correlations were found in all courses regardless of whether the skills were explicitly taught (WRIT 100 and WRIT 200), embedded (PFP 200 and PFP 211) or subsumed in CT theoretical examinations (GCRT 100). Taken together, we believe the above assessments of multiple forms of validity authenticate to a substantial degree the CTWC skills scorecard as valid. However, mindful that the effectiveness of the scorecard needs to be established not only by validity, but also through reliability, we turn to this analysis.

Reliability

What is the consistency of the CTWC skills scorecard when it is repeated? Is the scoring among faculty consistent (inter-rater reliability)?

Inter-rater reliability was assessed by determining the degree to which three faculty members provided consistent ratings of the CTWC skills scorecard items and totals across a subset of 24 randomly selected students' exams in the reading and writing course, WRIT 100.[27] Data analysis found that the IRR for total scores on the CTWC skills scorecard was excellent. Similarly, the IRR for total scores on comprehension and integration of writer's ideas was excellent, and excellent to good levels of IRR were also found in four of the six scorecard categories. This suggests that the variation in the CTWC skills ratings is due to student achievement rather than instructor error. However, there were two scorecard categories — inference and explanation — where the IRR were fair and poor respectively, suggesting some need for further training of faculty (Saxton et al., 2012). Accordingly, we argue that the CTWC skills scorecard can be used by faculty to score student work in a consistent manner, provided proper training is provided before assessment.

Do the individual items that make up the skills scorecard consistently measure the same underlying construct (internal consistency)?

26 In WRIT 100, 102 exams grades were assessed (out of the total 157). In GCRT 100, exam grades were evaluated for the total sample (137).
27 The ICC was calculated for each score on the individual items of the rubric as well as for the sums of total scores for comprehension, integration of writer's ideas and overall CTWC skills.

The reliability of the skills scorecard was further assessed for internal consistency to determine if the individual items that make up the scorecard consistently measure the same underlying construct.[28] Overall, internal consistency for all six CTWC skills items (total scales) in all courses was excellent. Therefore, we offer that the CTWC skills scorecard is highly reliable across all student samples in making stable and consistent measurements, which is in line with earlier high assessments of scale reliability in the pilot study (Humber College, 2013–2014, pp. 6, 18). And since reliability is a necessary condition for validity, this lends further weight to the argument that the CTWC skills scorecard is an effective assessment instrument.

8.2 Panel Study Results

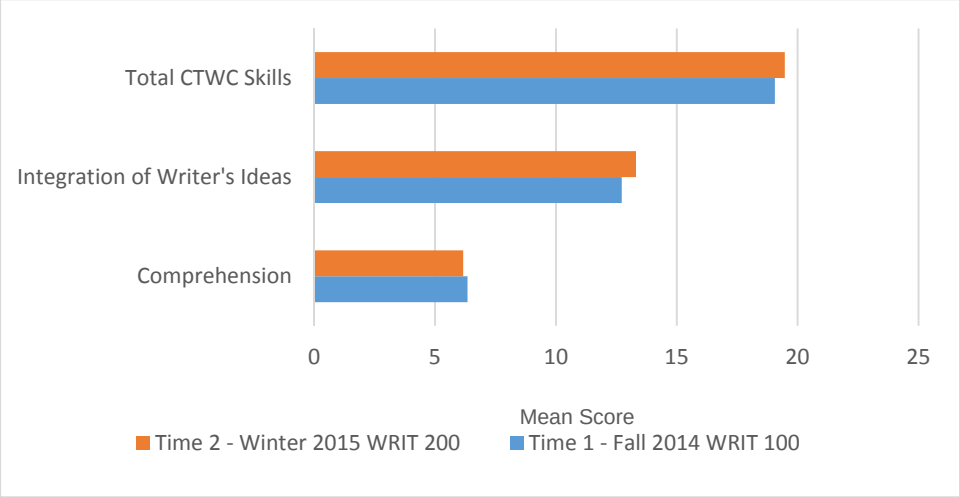
Presumably, over time and with focused attention, CTWC skills achievement will increase. Given that our research focused on two consecutive courses where CTWC skills were explicitly taught and assessed, we were able to evaluate the impact of the CTWC skills-building curriculum on students' levels of achievement over time, thereby posing a more fundamental question: Does the early, deliberate teaching of the essential skills result in gains in learning the outcomes (the longitudinal research design)?

To answer this question, the CTWC skills scorecard was used to assess the panel group's final exams from WRIT 100 and WRIT 200 in the Fall 2014 and Winter 2015 semesters. While limited in size and scope, the data analysis found students' average (mean) levels of achievement on the CTWC skills over the two semesters to be good on all individual items of the skills scorecard and good to very good on the total scores and summated scores for comprehension and integration of writer's ideas (Figure 6). Analysis of the descriptive statistics (means, standard deviations) indicates that students are typically learning the outcomes rather consistently and at higher levels (Appendix H).[29]

28 Cronbach's alpha was calculated for the CT and WC skills items and subscales (comprehension, integration of writer's ideas) in all courses. Internal reliability was also assessed through item-total correlations between sub-scores on individual items and overall total CTWC scores (Appendix H). Notably, the internal reliability (internal consistency) of the earlier rubric in the 2013–2014 data set was also examined through similar correlational analyses (Humber College, 2013–2014, pp. 6, 18).

29 As previously stated, a more detailed presentation of the data analysis and interpretation is available in Appendix H.

Figure 6: CTWC Skills Scores by Time (2014–2015)

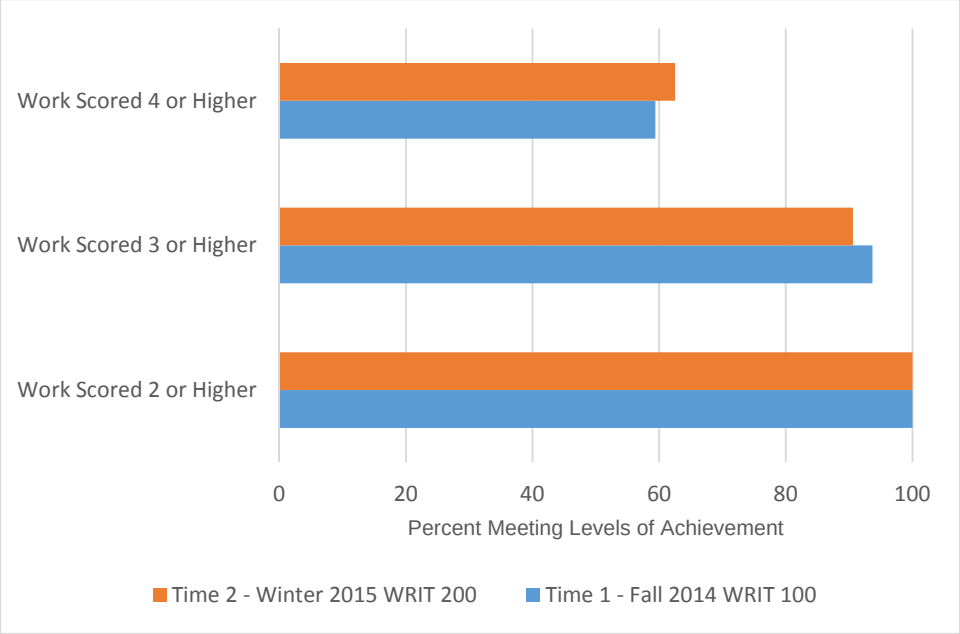


Note: Total CTWC skills scores range from 1–6 (poor), 7–12 (fair), 13–18 (good), 19–24 (very good) to 25–30 (excellent). Total comprehension scores range from 1–2 (poor), 3–4 (fair), 5–6 (good), 7–8 (very good) to 9–10 (excellent). Total integration of writer's ideas scores range from 1–4 (poor), 5–8 (fair), 9–12 (good), 13–16 (very good) to 17–20 (excellent).

However, means may not be the best option to track student growth (Greenhoot & Bernstein, 2011, p. 6). Instead, many scholars recommend additionally representing student learning assessment as distributions of performance across categories, that is, percentages of student work scored at or above the various milestone levels of achievement (Greenhoot & Bernstein, 2011, p. 6). When applied to our panel group, the results indicate that by the end of WRIT 200 in semester two, over 90% of students in the panel group achieved good or better total CTWC skills scores that aligned with learning outcome expectations (Figure 7). In fact, the vast majority achieved very good to excellent total CTWC skills scores, surpassing expectations. This said, gains seem to be happening more prominently with regard to the individual skills categories related to integrating writer's ideas (especially inference and explanation) rather than comprehension (interpretation and analysis).

We must also stress that the data analysis did not reveal any significant mean differences or gains in learning between students' skills scores over the two semesters we examined. This leaves more questions than answers, specifically regarding the deliberate teaching and mapping of CTWC skill building across the curriculum.

Figure 7: Total Scores on CTWC Skills by Time (2014–2015)



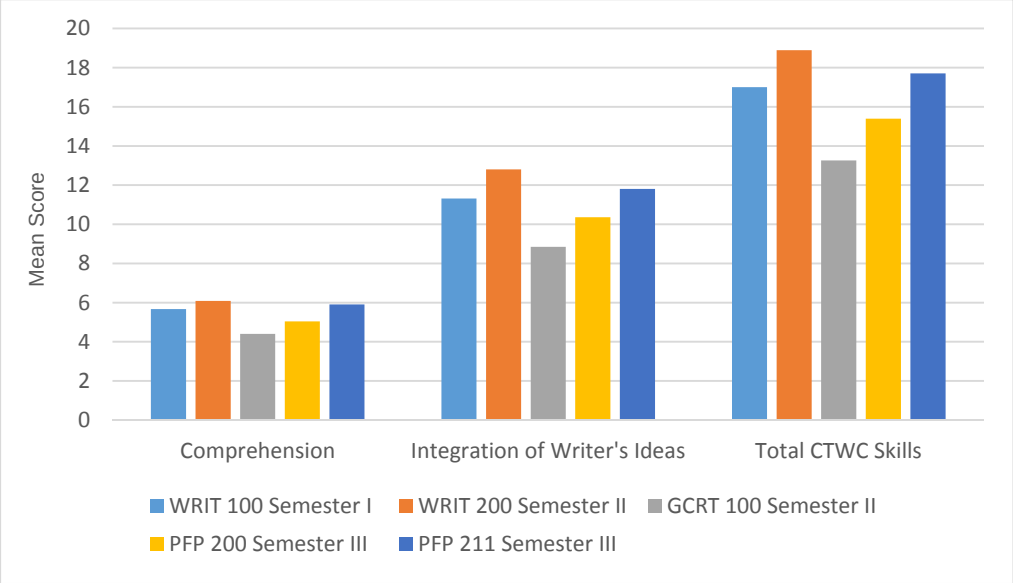
Note: Total CTWC skills scores range from: 1 (poor), 2 (fair), 3 (good), 4 (very good) to 5 (excellent).

8.3 Course/Program Comparisons

Although we did not find students are making significant gains in learning the outcomes over the duration of the cross-college reading and writing courses in which CTWC skills are explicitly taught, perhaps these students experiencing the early, special skills-building curriculum are still obtaining higher achievement levels than students in other learning contexts? Thus, the research question we examine next is: Across different student groups in different cross-college and program-specific courses, do diverse learning experiences in curricula related to CT and WC result in different levels of skill achievement (the cross-sectional research design)?

When analyzing the total CTWC skills scores by course (Figure 8) and related learning environment, we found that the early, explicit skills-building curriculum that student groups are experiencing in the diploma and advanced diploma cross-college reading and writing courses appears to have a more significant impact on their CTWC skills development. This is evidenced by these WRIT students' higher mean, and more consistent, scores on the total and summated CTWC learning outcomes (Table 2) (Appendix H).

Figure 8: CTWC Skills Scores by Course (2014–2016)



Note: Total CTWC skills scores range from 1–6 (poor), 7–12 (fair), 13–18 (good), 19–24 (very good) to 25–30 (excellent). Total comprehension scores range from 1–2 (poor), 3–4 (fair), 5–6 (good), 7–8 (very good) to 9–10 (excellent). Total integration of writer's ideas scores range from 1–4 (poor), 5–8 (fair), 9–12 (good), 13–16 (very good) to 17–20 (excellent).

In contrast, CTWC total skills scores achieved by students in the senior diploma PFP courses offering no overt teaching of the essential skills were lower on average and more variable (Table 2).

Early student learning experiences focused on building the essential CTWC skills rather than on covering conceptual and theoretical knowledge related to CT also appear to matter more. In other words, the ways in which the CT skills are taught seem to make some difference with regard to the levels of student success in achieving the outcomes. The GAS student group taking the content heavy CT course (GCRT 100) as part of their certificate program had, on average, the lowest total CTWC skills scores. Thus, even though the GAS students are the only ones in the college to take a special conceptual and theoretical CT course, this course content does not seem to automatically translate into better essential CTWC skills achievement. What matters more appears to be the explicit and applied approach in which the skills are taught.

Table 2: Differences in Total CTWC Skills by Course (2014–2016)

Course	Total Score on CTWC Skills			N
	M	MDN	SD	
WRIT 100	17.00	17.00	5.47	156
WRIT 200	18.89	20.00	5.12	120
PFP 200	15.39	15.00	6.90	256
PFP 211	17.71	19.00	6.23	53
GCRT 100	13.26	12.00	4.63	137

Note: M = Mean; MDN = Median; SD = Standard Deviation; N = Sample Size
Total CTWC skills scores range from 1–6 (poor), 7–12 (fair), 13–18 (good), 19–24 (very good) to 25–30 (excellent).

Overall, then, we suggest that this preliminary comparative evidence could mean that the early, explicit skills-building curriculum in the cross-college reading and writing courses makes a difference with regard to success in learning the outcomes. Further, when contrasted with the GAS group, whose theoretical, content-heavy learning experience in the CT course produced the lowest levels of student achievement, we argue that the positive impact of the cross-college reading and writing explicit skills-building journey on student skill attainment seems even more probable. This said, we certainly recognize the threats to internal and external validity inherent in the cross-sectional design of the study that prevent us from making any firm causal conclusions here or from generalizing the results to the larger Humber community. We further acknowledge the longitudinal results above that did not find significant gains in learning the skills after the cross-college reading and writing panel group's student work was assessed over two initial time periods. These limitations of research design and the study, along with related suggestions for future research to better ascertain the effect of curricula, student learning and skill achievement over the entire span of students' PSE programs, are discussed in a later segment. More than anything, it is imperative that we also engage faculty in any such discussions as the success of any assessment tool relies fully on its adoption by faculty. This then is the focus of the last part of this study.

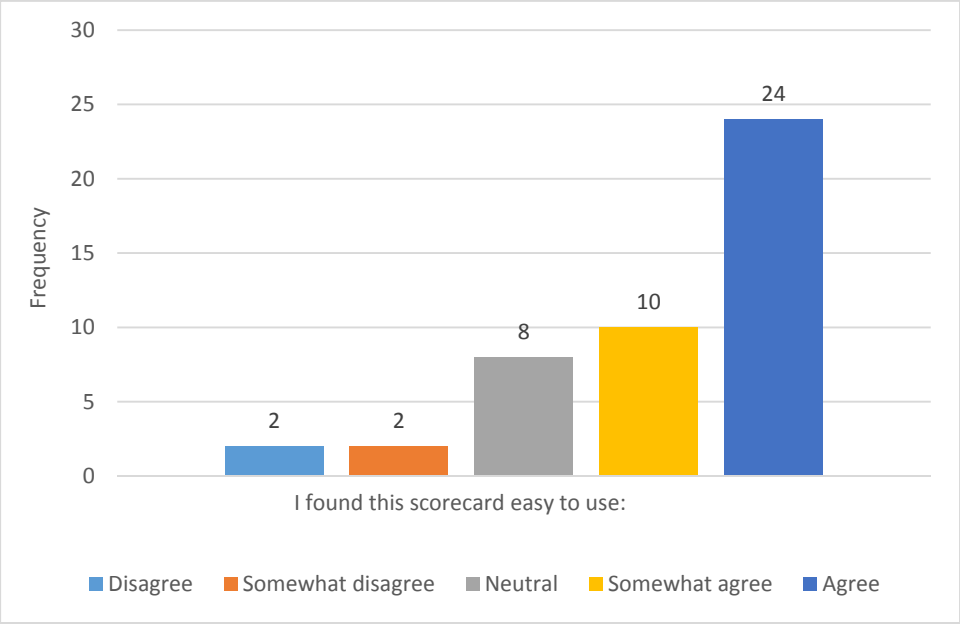
8.4 Usability and Scalability Results

For the CTWC skills scorecard to be implemented and scaled successfully across the institution, it is important that faculty find it both easy to use and able to supply important information to students about their CTWC skills attainment and any progress required. For example, it is essential that the scorecard be simple enough for faculty to apply consistently and reliably and to experience the overall scoring process as simple and straightforward. It is also imperative that faculty perceive the scorecard to be valid and capable of supplying students with significant signals about the skills learned so that they see the benefit of the assessment and are invested in the project from the outset. Hence, we ask several questions. Do faculty find the scorecard easy to use? Do faculty believe the rubric will supply meaningful information to their students

about their level(s) of achievement in the CTWC areas? Can the CTWC skills scorecard be implemented optimally within the institution and scaled successfully?

In total, 46 faculty used the scorecard to assess their students' CTWC skills and 34 (74%) agreed that the scorecard was easy to use (Figure 9). This bodes well for the eventual scaling of the tool across the institution.

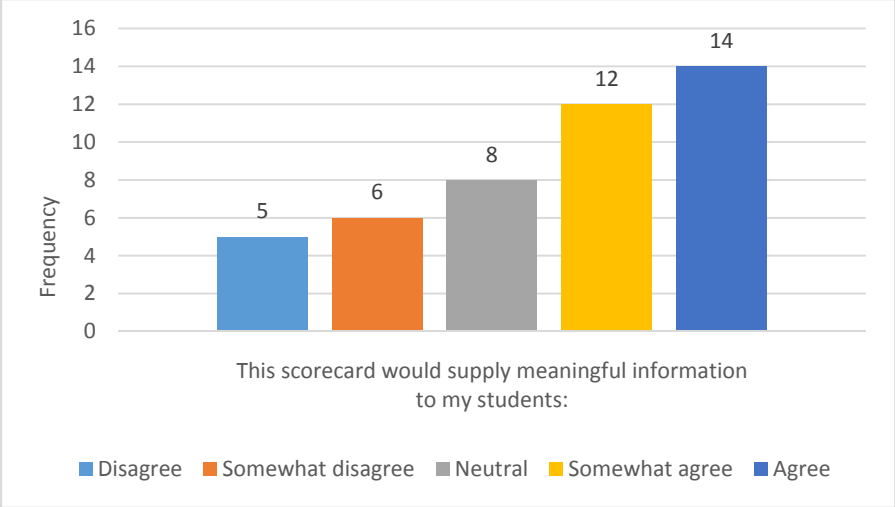
Figure 9: Faculty Perceptions of the Usability of the CTWC Skills Scorecard (2013–2016)



However, when asked if the scorecard would supply meaningful information to students about the skills learned, only 26 (58%) of faculty agreed that it would (Figure 10).[30]

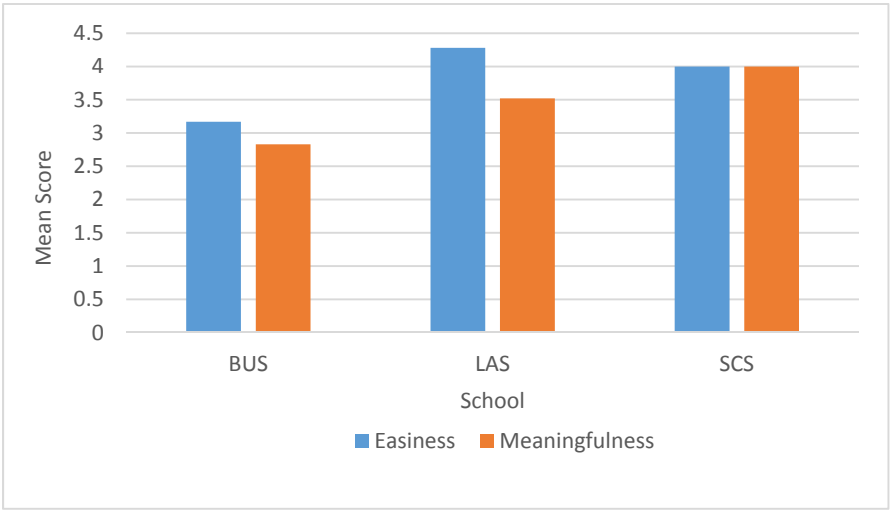
30 During the study, students were not provided with any feedback about the CT assessment. The CT assessment was completed by faculty as a supplemental grading of students' work and strictly for research purposes.

Figure 10 Faculty Perceptions of the Meaningfulness of the CTWC Skills Scorecard (2013–2016)



Further, although faculty generally agreed overall that the scorecard was straightforward and that the information it provided was somewhat significant, this was not consistent across faculty from different schools (Figure 11). In particular, faculty from LAS had higher agreement, on average, that the assessment tool was user-friendly when compared with faculty from BUS and SCS.

Figure 11: Faculty Perceptions of the Usability and Meaningfulness of the CTWC Skills Scorecard by School (2013–2016)



Note: Faculty perceptions of the usability and meaningfulness of the skills scorecard were measured on a numeric scale ranging from 1 (disagree), 2 (somewhat disagree), 3 (neutral), 4 (somewhat agree) to 5 (agree).

This suggests that faculty who teach in areas where CTWC skills are more explicitly taught are more familiar with the ways in which these skills can be meaningfully assessed. As such, in order to successfully scale the tool, significant investment in faculty training is needed to ensure that all faculty, regardless of their background or school affiliation, are prepared to effectively use the tool.

9. Discussion

We tentatively hypothesized that the early, explicit CTWC skills-building student experience in the two cross-college reading and writing courses would result in gains in learning the skills over the two semesters as well as in comparatively overall higher levels of achievement than in the alternative, more implicit approaches to teaching the skills in the PFP-specific courses, and in the GAS content-heavy CT course. In the first case, we examined the relationship between the explicit skills-building curriculum and students' CTWC skills scores over time. In the second case, we explored the relationship between the different curricula and the associated diverse student learning experiences in the selected courses and different levels of achievement on students' CTWC skills scores.

Upon analysis of the data, we could not validate the initial hypothesis that explicit CTWC skills building will result in gains in learning the skills. However, when the CTWC skills of students exposed to explicit skill building (WRIT 100 and 200) were compared with the gains of students in courses where the skills are more implicitly embedded (PFP 200 and PFP 211) or in a discrete course focused on CT as a discipline (GCRT 100), we found that the WRIT 100 and 200 students had comparatively higher levels of achievement. This would suggest that CTWC skills need to be taught explicitly rather than implicitly. Furthermore, discrete courses on CT theory are not necessarily the most effective means by which to develop, acquire and practice CT skills.

Further, CTWC skills, like any other skill, need to be developed and practised over time. As stated by Joordens, Paré and Collimore (2014, p. 6), "[o]ne cannot be told how to think critically or how to write well. Students must practice the skill repeatedly within a context that provides structure and guidance." Our data analysis suggests that in the context of Humber College, this occurs in the mandatory reading and writing courses that all certificate and diploma students must take. However, these skills cannot be mastered by all in a two-semester period, as demonstrated by the initial empirical evidence. Although our panel study included only two student products in two courses (WRIT 100 and WRIT 200), and two periods of time to compare, based on the findings of limited gains in learning we propose that CTWC skills need more time to mature. The development of these skills can be further enhanced by courses where these skills are embedded, but students benefit most from explicit skills building over a more extended period of time. Consequently, we would recommend that explicit CTWC skills-building courses be positioned in each year of a student's program of study to allow for sufficient time for these skills to develop and mature.

This leads us to the final question that we asked in this study regarding the usability and hence the scalability of the skills scorecard. The revision to the skills scorecard between the pilot phase and the second phase of this study indicates that usability and scalability were of the utmost importance. Our data analysis indicates that the revised scorecard was user-friendly, even though faculty were not as convinced that the

scorecard would provide meaningful information for students. However, our reliability analysis also provided additional information that would be important to consider when discussing scalability. First, the reliability results suggest that attention can and should be paid to rater selection and training to increase raters' consistent judgements on items of the rubric such as inference and explanation. Research has shown that rater training can improve agreement (accuracy), reliability and validity, emphasizing the need for raters to develop confidence and a common understanding when using the assessment tool in order to apply it as consistently as possible (Hoyt & Kerns, 1999). Research also indicates that selecting expert raters can further improve accuracy and reliability as raters with expertise in the related fields of the scorecard will be better able to differentiate between the varying domains of the performance scale (Smither, Barry & Reilly, 1989 as cited in Graham et al., 2012, p. 16). Indeed, it is essential for instructors to understand the individual items or domains of the scorecard to improve intersubjective understanding and inter-rater agreement on all levels of CTWC skills achieved by students. This need for qualified and trained raters further supports our suggestion that the assessment of CTWC skills should occur in courses where these skills are taught explicitly, and where faculty are already familiar with the concepts of CTWC, the language used to assess CTWC skills and the achievement levels of students in CTWC.

10. Limitations, Suggestions for Future Research and Overall Recommendations

To summarize, the CTWC skills scorecard developed and piloted by Humber College in the 2013–2014 academic year, and then revised and applied across several courses and programs in the 2014–2016 academic years (assessing essential skills achievement in courses where these skills are explicitly taught, where they are implicitly embedded and where they are examined as a discipline), was found overall to be both a valid and reliable tool. In addition, descriptive statistics from the course comparisons in the second phase of the study demonstrated that most students in the cross-college reading and writing courses (WRIT 100 and WRIT 200), experiencing the new, explicit skills-building curriculum, are learning overall CTWC skills at good or superior levels and typically have higher total skills scores than student groups from other diploma courses (PFP and certificate GAS), where the essential skills are not explicitly taught.

Thus, overall, we suggest that this preliminary comparative evidence could mean that the new, explicit skills-building curriculum that students are experiencing in the cross-college reading and writing courses makes a difference with regard to success in attaining the essential skills outcomes. This seems even more probable when contrasted with the GAS group, whose theoretically, content-heavy learning experience in the CT course produced the lowest levels of student achievement. This said, we certainly recognize the threats to internal and external validity inherent in the cross-sectional design of the study, which prevented us from drawing any firm causal conclusions or generalizing the results to the larger Humber community. We further acknowledge the longitudinal results that did not find significant gains in learning the skills after the cross-college reading and writing panel group's student work was assessed over two initial time periods. As a result, we would suggest that future research be conducted to better ascertain the effect of curricula, student learning and skill achievement over the entire span of students' PSE programs. To this end, we

would suggest that a new panel study be undertaken to assess CTWC gains over time that increases the number of student products and the time periods being compared.

To further enhance the comparative study undertaken in this research, we would suggest that the longitudinal study be expanded to include panel comparisons using programs of the same duration, hence reflecting the same qualifications under the Ontario Qualification Framework. Such programs should also be drawn from a range of industries where, ideally, CTWC skills are in high demand. This would allow for a more robust study to be undertaken of the progressive achievement levels of students in CTWC skills building, which would, in turn, strengthen the usability of the tool.

As indicated from our data analysis, faculty generally and quite consistently agreed that the skills scorecard is user-friendly. We believe these findings speak rather positively to the scorecard's prospects for implementation across the institution and for successful scaling. If we focus the question of scalability solely on the ability of faculty to reliably use the skills scorecard, then, with a dedicated focus on rater recruitment and training, we can confidently proceed with scaling the CTWC skills scorecard across the institution, providing we focus on courses where CTWC skills are explicitly taught. In the case of Humber College, these would be the mandatory reading and writing courses found in every certificate and diploma program.

However, the scalability of the skills scorecard ultimately hinges on its ability to provide meaningful information to students about their skill levels and, ideally, assist them in tracking gains in their skills over time. Here, this study revealed less agreement and slightly more variable opinion among faculty as to whether the tool would allow students to do this in a meaningful manner. We would suggest that to build greater meaningfulness into the tool for students, a common understanding of CT, including the cognitive and metacognitive skills associated with CT, needs to be established among faculty across programs. Indeed, a more intentional and consistent approach to how CT is taught (whether explicitly or implicitly), referenced, reinforced and assessed across programs and curricula would establish a coherent framework within which students could better understand the essential skills and their attainment levels. Further, presenting a more direct and consistent articulation for students of how the essential skills relate to their future employability would further build meaningfulness into the skills scorecard by allowing students to harness their attainment levels on CTWC when presenting their candidacy to employers.

In the absence of an integrated approach of this kind, which engages all stakeholders, the CTWC scorecard will likely have little value for students beyond the scores that they achieve, and even less value for employers — a lost opportunity indeed, as employers consistently demand that colleges produce work-ready graduates with the knowledge, skills and abilities needed to contribute in a meaningful way to our knowledge-based, global economy. Thus, a common understanding of the value in assessing essential skills beyond the obvious need for program accountability and quality is a prerequisite for developing, implementing and supporting both a user-friendly and meaningful assessment tool, such as the CTWC scorecard.

11. List of Abbreviations

APA	American Philosophical Association
AHELO	Assessment of Higher Education Learning Outcomes
AACU	Association of American Colleges and Universities
BACC 300	Financial Analysis and Accounting Course
BFIN 420	Business Finance Course
BFIN 500	Corporate Finance Course
BMGT 300	Human Resource Administration Course
BMGT 500	Strategic Management Course
BUS	School of Business
CCTST	California Critical Thinking Skills Test
CLA	Collegiate Learning Assessment
CSAC	College Standards and Accreditations Council
CI	Confidence Interval
COMM 100	Introduction to College Reading and Writing Skills Remedial Course (now WRIT 050)
COMM 200	College Reading and Writing Skills Course (now WRIT 100)
COMM 300	Workplace Writing Skills Course (now WRIT 200)
CT	Critical Thinking
CTWC	Critical Thinking and Written Communication
df	Degrees of Freedom
EES	Essential Employability Skills
ESL	English as a Second Language
GAS	General Arts and Sciences
GCRT 100	Critical Thinking Course
HEQCO	Higher Education Quality Council of Ontario
ICC	Intraclass Correlation Coefficient
IRR	Inter-rater Reliability
KPI	(Graduate and Employer) Key Performance Indicators
LOAC	Learning Outcomes Assessment Consortium
LAS	School of Liberal Arts and Sciences
MDN	Median
MSIT	School of Media Studies and Information Technology
M	Mean
MAESD	(Ontario) Ministry of Advanced Education and Skills Development (formerly MTCU)
MTCU	(Ontario) Ministry of Training, Colleges and Universities (now MAESD)
N	Sample Size
OECD	Organization for Economic Co-operation and Development
OBE	Outcomes-based Education
PFP	Police Foundations Program
PFP 200	Criminal and Civil Law Course

PFP 211	Interviewing and Investigations Course
PFP 301	Criminal and Civil Law Course
PFP 403	Police Oversight Course
PSE	Post-Secondary Education
PIAAC	Programme for the International Assessment of Adult Competencies
PISA	Programme for International Student Assessment
SCAPA	School of Creative and Performing Arts
SCS	School of Social and Community Services
SD	Standard Deviation
SPSS	Statistical Package for the Social Sciences (IBM)
TER	Test of Everyday Reasoning
VALUE	Valid Assessment of Learning in Undergraduate Education
WC	Written Communication
WRIT 100	College Reading and Writing Skills Course (formerly COMM 200)
WRIT 200	Workplace Writing Skills Course (formerly COMM 300)
WRIT 050	Introduction to College Writing Skills Remedial Course (formerly COMM 100)

References

Arum, R., & Roksa, J. (2010). *Academically adrift: Limited learning on college campuses.* Chicago, IL: University of Chicago Press.

Benjamin, R. (2013). Three principle questions about critical-thinking tests. New York: Council for Aid to Education.

Benjamin, R., Klein, S., Steedle, J., Zahner, D., Elliot, S., & Patterson, J. (2013). The case for critical thinking skills and performance assessment. New York: Council for Aid to Education.

Bok, D. (2006). *Our underachieving college: A candid look at how much students learn and why they should be learning more.* Princeton, NJ: Princeton University Press.

Brochu, P., Deussing, M-A., Houme, K., & Chuy, M. (2013). *Measuring up: Canadian results of the OECD PISA Study. The performance of Canada's youth in mathematics, reading and science.* Canada: Council of Ministers of Education.

Bryman, A., Teevan, J. J., & Bell, E. (2009). *Social research methods* (2nd Canadian ed.). Don Mills, ON: Oxford University Press.

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada. (2014). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.* Retrieved from http://www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS_2_FINAL_Web.pdf

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290.

Ciudad-Gómez, A., & Valverde-Berrocoso, J. (2014). Reliability analysis of an evaluation rubric for university accounting students: A learning activity about database use. *Journal of International Education Research*, *10*(5), 301–307.

Cohen, J. W. (1988). *Statistical power analysis for the behavioural sciences*. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Colleges Ontario. (2016). *Key Performance Indicators 2015.*

Colleges Ontario. (2015). *Key Performance Indicators 2014.*

Dawson, D., Borin, P., Meadows, K., Britnell, J., Olsen, K., & McIntyre, G. (2014). *The impact of the instructional skill workshop on faculty approaches to teaching.* Toronto: Higher Education Quality Council of Ontario.

Deller, F., Brumwell, S., & MacFarlane, A. (2015). *The language of learning outcomes: Definitions and assessments.* Toronto: Higher Education Quality Council of Ontario.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: Sage.

Facione, P. (2013). *Critical thinking: What it is and why it counts.* Retrieved from https://www.insightassessment.com/Resources/Importance-of-Critical-Thinking/Critical-Thinking-What-It-Is-and-Why-It-Counts

Finley, A. (2011). How reliable are the VALUE rubrics? *Peer Review, 13/14*(4/1), 1–6.

Fisher, R., & Hoth, W. (2010). *College-level literacy: An inventory of current practices at Ontario's Colleges.* Toronto: Higher Education Quality Council of Ontario.

Foundation for Critical Thinking. (2015). Critical thinking: Where to begin. Retrieved from http://www.criticalthinking.org/pages/critical-thinking-where-to-begin/796

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn and Bacon.

Goff, L., *et al.,* (2015). *Learning outcomes assessment: A practitioner's handbook*. Toronto: Higher Education Quality Council of Ontario.

Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. Center for Educator and Compensation Reform, U.S. Department of Education. 34705.0212.83670507.

Greenhoot, A., & Bernstein, D. (2011). Using VALUE rubrics to evaluate collaborative course design. *Peer Review, 13/14*(4/1), 1–8.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quantitative Methods Psychology, 8*(1), 23–24.

Hansson, E. E., Svensson, P. J., Strandberg, E. L., Troein, M., & Beckman, A. (2014). Inter-rater reliability and the agreement of rubrics for assessment of scientific writing. *Education, 4*(1), 12–17.

Hays, R. D., & Reviki, D. A. (2005). Reliability and validity (including responsiveness). In P. M. Fayers & R. D. Hays (Eds.). *Assessing quality of life in clinical trials: Methods and practice* (25–39). New York, NY: Oxford University Press.

Hoyt, W. T., & Kerns, M. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods, 4*(4), 403–424.

Humber College (2014, May 23). *HEQCO: LOAC, Steering and Pilot Groups* (meeting notes). Toronto.

Humber College (2014). *Generic Skills Assessment Study – Semester 1 Findings.* Toronto.

Humber College (2013–2014). *Employability skills assessment*. Toronto.

Humber College (2013). *Strengthen, Sustain, Maximize. 2013–2018 Strategic Plan*. Toronto: Humber Strategic Planning and Institutional Analysis, Humber College.

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

Joordens, S., Paré, D., & Collimore, L-M. (2014). *Taking learning outcomes to the gym: An assignment-based approach to developing and assessing learning outcomes.* Toronto: Higher Education Quality Council of Ontario.

Kaupp, J., Frank, B., & Chen, A. (2014). *Evaluating critical thinking and problem solving in large classes: Model eliciting activities for critical thinking development.* Toronto: Higher Education Quality Council of Ontario.

Landers, R. N. (2015). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *The Winnower*. *2*:e143518.81744. DOI: 10.15200/winn.143518.81744.

Lennon, M. C. (2014). *Piloting the CLA in Ontario.* Toronto: Higher Education Quality Council of Ontario.

Lennon, M. C. (2010). *Signalling abilities and achievement: Measuring and reporting on skill and competency development.* Toronto: Higher Education Quality Council of Ontario.

Lennon, M. C. (2010). *A fine balance: Supporting skills and competency development.* Toronto: Higher Education Quality Council of Ontario.

Lennon, M. C., & Jonker, L. (2014). *AHELO: The Ontario Experience.* Toronto: Higher Education Quality Council of Ontario.

Lennon, M. C., Frank, B., Humphreys, J., Lenton, R., Madsen, K., Omri, A., & Turner, R. (2014). *Tuning: Identifying and measuring sector-based learning outcomes in postsecondary education.* Toronto: Higher Education Quality Council of Ontario.

Levin, J., Fox, J. A., & Forde, D. R. (2010). *Elementary statistics in social research* (11th ed.). Boston, MA: Pearson Education Inc., publishing as Allyn and Bacon.

Liu, Q. (2015). *Outcomes-based education initiatives in Ontario postsecondary education: Case studies*. Toronto: Higher Education Quality Council of Ontario.

Martini, T. S., & Clare, M. (2014). *Undergraduates' understanding of skill-based learning outcomes: Can e-portfolios help?* Toronto: Higher Education Quality Council of Ontario.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46.

Mora, M. (2016). How the Source of Your Data Impacts Your Survey Results. *SurveyGizmo.*

Ontario Ministry of Advanced Education and Skills Development (formerly MTCU). (2003). *Minister's binding policy directive: Framework for programs of instruction*. Retrieved from http://www.tcu.gov.on.ca/pepg/documents/FrameworkforPrograms.pdf

Ontario Ministry of Advanced Education and Skills Development. (2009a). *Ontario qualifications framework.* Toronto: Ministry of Training, Colleges and Universities. Retrieved from http://www.tcu.gov.on.ca/pepg/programs/oqf/QsAsOQF.html

Ontario Ministry of Advanced Education and Skills Development. (2009b). *Ontario essential employability skills*. Toronto: Ministry of Training, Colleges and Universities. Retrieved from http://www.tcu.gov.on.ca/pepg/audiences/colleges/progstan/essential.html

Ontario Ministry of Advanced Education and Skills Development. (2009b). *Ontario program standards*. Toronto: Ministry of Training, Colleges and Universities. Retrieved from http://www.tcu.gov.on.ca/pepg/audiences/colleges/progstan/history.html

Organization for Economic Co-operation and Development (2015). Canada, *Education at a Glance 2015: OECD Indicators.* Paris: OECD Publishing. Retrieved from http://dx.doi.org/10.1787/eag-2015-47-en

Pallant, J. (2013). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (5th ed.). Berkshire, England: Open University Press, McGraw-Hill Education.

Pascarella, E. T., and Terenzini, P. T. (2005). *How college affects students. A third decade of research.* San Francico, CA: Jossey-Bass.

Pusecker, K. L., Torres, M. R., Crawford, I., Levia, D., Lehman, D., & Copic, G. (2011). Increasing the validity of outcomes assessment. *Peer Review, 13/14*(4/1), 1–8.

Saxton, E., Belanger, S., & Becker, W. (2012). The Critical Thinking Analytic Rubric (CTAR): Investigating intra-rater and inter-rater reliability of a scoring mechanism for critical thinking performance assessments. *Assessing Writing, 17,* 251–270.

Scriven, M., & Paul, R. (1987). *Defining critical thinking*. Retrieved from http://www.criticalthinking.org/pages/defining-critical-thinking/766

Siefert, L. (2011). Assessing general education learning outcomes. *Peer Review, 13/14*(4/1), 1–7.

Smither, J. W., Barry, S. R., & Reilly, R. R. (1989). An investigation of the validity of expert true score estimates in appraisal research. *Journal of Applied Psychology, 74,* 599–605.

Statistics Canada. (2013). *Skills in Canada: First results from the Programme for the International Assessment of Adult Competencies (PIAAC).* Statistics Canada. Catalogue no.89-555-X. Retrieved from http://www.statcan.gc.ca/pub/89-555-x/89-555-x2013001-eng.pdf

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

Velyvis, V.P., Suen, W., & Nguyen, R. (2016). *Selecting measures*. Research Education Series: Module III. Inter-Professional Research, Knowledge Translation and Academic Development. Ontario Shores Centre for Mental Health Sciences.

Weingarten, H. (2014, January 21). Learning outcomes: The game changer in higher education. It's not academic. *Thoughts, ideas, kudos & brickbats from HEQCO*. Retrieved from http://blog-en.heqco.ca/2014/01/harvey-p-weingarten-learning-outcomes-the-game-changer-in-higher-education/

Weingarten, H. (2014, February 13). Managing for quality. Classifying learning outcomes. It's not academic. *Thoughts, ideas, kudos & brickbats from HEQCO.* Retrieved from http://blog-en.heqco.ca/2014/02/harvey-p-weingarten-managing-for-quality-classifying-learning-outcomes/

Higher Education Quality Council of Ontario

An agency of the Government of Ontario