Implementation Evaluation of a Critical Online Resource Evaluation (CORE) Program for High School Students in Quebec

Maria Cecilia Jimenez Fernandez

A Thesis in

The Department of

Education

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Arts (Education) at

Concordia University

Montréal, Quebec, Canada

August 2023

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Maria Cecilia Jimenez Fernandez

Entitled: Implementation Evaluation of a Critical Online Resource Evaluation (CORE) Program for High School Students in Quebec

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Educational Technology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair
Dr. Julie Corrigan

_____ Examiner
Dr. Steven Shaw

_____ Examiner
Dr. Ji Yae Bong

_____ Thesis Supervisor
Dr. Julie Corrigan

Approved by _____
Dr. Sandra Martin-Chang, Graduate Program Director

Month day (Date of defence), 2023

_____
Dr. Pascale Sicotte , Dean of Faculty

Abstract

Implementation Evaluation of a Critical Online Resource Evaluation (CORE) Program for High School Students in Quebec

Maria Cecilia Jimenez Fernandez

Being literate in the 21st century goes beyond the abilities of reading and writing; new skills and methods are required to evaluate the credibility of the information we find online. It is frequently assumed that young students, the so-called 'digital born generation,' can discern credible information from mis/dis information or biased information; however, evidence suggests they have limited critical evaluation skills. Aware of this gap, researchers from the University of Concordia and practitioners in three high schools in Quebec implemented a Critical Online Resource Evaluation (CORE) intervention program to improve students' ability to evaluate online resources. With the vision to scale up the intervention in Canada and worldwide, it was essential to understand how the program was implemented and whether it was delivered as intended by the research team, to make informed decisions about modifications, adaptations, or refinements to improve the program's effectiveness.

Implementation fidelity was evaluated using a pragmatic mixed methods design, concurrently collecting and analyzing quantitative and qualitative data from three sources: teachers, researchers and students. The CORE intervention was implemented with high fidelity from both teacher and researcher perspectives (79.9%, 89.0%). While teachers adhered to the program content as planned, they adapted it to meet their classroom context. Teachers conducted the program with a high level of quality, with a mean score of 90.3% from the teacher's perspective and 96.7% from the researcher's perspective, suggesting they were well-prepared and confident. Students' engagement and motivation varied from the different views. From the teacher's perspective, students were highly engaged, with a mean score of 80.9%, while researchers and students perceived they were moderately engaged and motivated (77%,

69.6%). Teachers, researchers, and students concurred that the program was long, complex and the COVID topic fatiguing. Finally, the intervention highlighted group work and interactivity as the most engaging and motivating factors.

Acknowledgements

Since I was studying for my bachelor's in Systems Engineering in the 1980s, I dreamed of doing my master's degree. However, my life path deviated a 'bit' from that dream. I became a devoted mom of four kids, a dedicated wife, and a committed professional worker. My desire to continue my studies became a low priority, but it never disappeared. Today at 59, I am close to achieving my dream and getting my MA in Educational Technology.

This dream fulfillment would not be possible without the collaboration of many people. First, a big thank you to my supervisor Dr. Julie Corrigan who played a vital role in my academic and research journey. She not only was an excellent teacher but generously shared a part of her project, hired me as RA, and allowed me to participate in the design and implementation of the CORE intervention program. She guided me from A to Z while developing my research, making my graduate student experience positive and productive. Dr. Steven Shaw, thank you for the enjoyable classes and for sharing a wealth of knowledge with me. Your deep understanding and expertise in the field will remain in my memory. Dr. Ji Yae Bong, even though it was a very short interaction during my attempt to take a course with you (unfortunately, I withdrew due to a personal unforeseen circumstance), it was enough to perceive the great teacher and person you are. Thank you all for accepting to be part of my evaluation committee.

Finally, I would like to express my heartfelt gratitude to my family. My wonderful husband Juan, your love and encouragement gave me the emotional support I needed to stay determined and committed to completing this master's and this thesis. Your willingness to take on additional responsibilities at work and home has enabled me to give my best to this journey. To my amazing kids, your love, smiles, and understanding during times when I had to be away with my work meant the world to me. Your presence has constantly reminded me of the purpose behind my efforts.

Dedication

To my beloved husband Juan and his unconditional accompaniment, to my children Mauricio, Sebastián, Andrés, and Isabella for believing in me, and to my dear grandson Samuel who arrived at the time of writing this thesis to give me energy and love.

A mi amado esposo Juan y su acompañamiento incondicional, a mis hijos Mauricio, Sebastián, Andrés e Isabella por creer en mí, y a mi adorado nieto Samuel que llegó en el momento de escribir esta tesis para llenarme de energía y amor.

Table of Contents

List of Figures

List of Tables

Chapter One: Introduction

Exposure to misinformation and disinformation increases as young students become more dependent and reliant on the internet and social media to get informed. Students can easily access over 200 million active websites (Fleming, 2019) and connect to billions of Tik Tok, Instagram, and Snapchat users worldwide[1]. In Canada, social network users stand at 34.47 million, with young people accounting for about 20% of the total share of users (Dixon, 2022). Through these social networks, young people access the news, get informed, get involved, replicate, and share a variety of information. The problem is that the difference between information and mis/dis information is not evident to them, leading to the potential spread of incorrect, misleading, or false information (Corrigan, 2019; Kiili et al., 2008; Kiili et al., 2017).

New skills and strategies to evaluate the credibility of online information are essential to prepare students for academic and personal success in the 21st Century (Leu et al., 2004). It is frequently assumed that these students born in the digital era possess these skills innately; however, evidence suggests that young students have limited critical evaluation skills to distinguish between reliable from unreliable information (Coiro et al., 2015; Corrigan, 2019; Kiili et al., 2008; Macedo-Rouet et al., 2019). Emerging empirical literature indicates that when properly developed and implemented intervention programs within the school setting for teaching students how to evaluate online resources' credibility critically can produce positive effects on students' skills (Brand-Gruwel et al., 2005; Hämäläinen et al., 2020; McGrew, 2020; Pérez et al., 2018; Zhang & Duke, 2011). Undertaking such programs in the classroom environment inherently entails evaluating whether such an intervention is working. Educational professionals and practitioners must evaluate the effects of the actions and programs they implement. If the intervention is not working as expected, it is necessary to promptly identify

[1] In 2022, the mobile app Tik Tok was downloaded 672 million times, followed by Instagram, with 548 million downloads. In the same year, Snapchat reached 330 million downloads (Ceci, 2023).

deficiencies and problems, establish a corrective action plan, and continue to monitor it. One critical measurement of implementation outcomes is *program fidelity*, which answers the question, was the program implemented as designed? Some authors refer to the same concept as treatment fidelity, treatment integrity, or fidelity of implementation (Breitenstein et al., 2010). Research on program fidelity for interventions in digital literacy is scarce worldwide (O'Donnell, 2008). Implementation fidelity is a relatively new construct in K-12 intervention programs, but it is not new in the evaluation practice (O'Donnell, 2008; Mowbray et al., 2003.) Moreover, it is now expected to be a component of the quality of any program evaluation.

Aware of the significance of evaluating school intervention programs, a group of researchers from Concordia University in Montreal[2] decided to evaluate the fidelity of implementing their research project entitled Critical Online Resource Evaluation (CORE) intervention program. The program was funded by the Ministère de l'Éducation et de l'Enseignement supérieur (MEES) and the Fonds de recherche du Québec – Société et culture (FRQSC), and it was implemented by several practitioners across three secondary educational institutions in Quebec. The researchers envision that by increasing the ability to evaluate the online resources that students consult or find on a daily basis, they can succeed academically and ultimately make better-informed decisions regarding their political, social, and economic lives. With this idea in mind, the research team leaders plan to expand their project to a broader population, and that is why it is imperative to evaluate not only the intervention outcomes (i.e., the ability to evaluate online resources) but also the implementation outcome (i.e., program fidelity), so they can strengthen the quality of the CORE program for future implementations. Measuring program fidelity will allow researchers and practitioners to understand how and why the intervention worked and the extent to which results can be improved through actions and

[2] Dr. Julie Corrigan (Principal Investigator), Assistant Professor of Digital Literacies at the University of Concordia; Dr. Kim McDonough, Professor and Canada Research Chair in Applied Linguistics at Concordia, and Dr. Heike Neumann, Senior Lecturer in English as a Second Language at Concordia.

decisions about the program (Carroll et al., 2007). For example, they can enhance the instructional design to ensure it is more responsive to student's needs and preferences; make pre-training for teachers more suitable; or improve student artifacts (e.g., workbooks, forms, questionnaires).

The present study's purpose is to evaluate the extent to which the CORE intervention program was delivered with fidelity in the context of five classrooms in three schools in Quebec. More specifically, this study evaluated three dimensions of program fidelity: (1) adherence to content (i.e., whether teachers adhered to the content to be delivered), (2) quality of delivery (i.e., whether teachers were prepared, and confident to deliver the program.), and (3) participant's responsiveness (i.e., whether the students were engaged and motivated) The proposed study includes the voices of teachers, students, and researchers in both quantitative and qualitative data formats. These multiple perspectives and formats led to analysis and discussions that may shed some insights on the implementation experience and inform the primary stakeholders on how to improve the program's quality and effectiveness and make informed decisions about future iterations of the intervention. Finally, the research question regarding whether and to what extent the students' CORE skills improved because of the intervention will be addressed in a separate study conducted by the principal investigator and the research team. However, I will reference these results in this study as they complement my own research questions.

Chapter Two: Background and Literature Review

This literature section examines Critical Online Resource Evaluation (CORE) as the intervention's subject, and it reviews the literature on program implementation fidelity, offering a foundation for the current study. First, the literature review addresses a larger area of new (digital) literacy intervention programs, narrowing down to credibility evaluation. Next, an examination of how to evaluate programs is presented, including fundamental definitions**,** approaches and standards and their importance during the development and implementation of educational interventions. A specific review focusing on implementation fidelity is presented as it is pertinent to the investigation. This review ends with the proposed current study and context.

The Subject of the Intervention: New (Digital) Literacies Intervention Programs

The Internet and Information and Communication Technologies (ICTs) have permeated almost all the spaces of our day-to-day lives, whether related to our professional, student, work or personal life and creating a more rapid pace of changes like literacy than previously seen in history. As Fraillon et al. (2015) stated in a worldwide study in which more than 20 countries participated, educational systems and educators worldwide recognize the importance of equipping students with tools and skills that enable them to participate in the digital era. The Quebec government is no stranger to this reality and has explicitly expressed its commitment to support the development of the digital skills of young people and adults in its "Digital Action Plan for Education and Higher Education" policy document (Ministère de l'Éducation et de l'Enseignement Supérieur, 2018). The CORE program was born in response to a call from the Government of Quebec to improve digital literacies within its educational system. The program is situated within the new literacies of online research and comprehension theory developed by Leu et al. (2017). This theory views online research as a reading skill involving online and offline reading. It further views online reading as an online inquiry process that encompasses identifying a problem, locating information, critically evaluating, synthesizing, and communicating information. Typically, new literacies build upon foundational literacies rather

than replace them altogether. However, these literacies alone are insufficient if we want to develop students' full potential by incorporating the Internet and ICTs. Comprehensive research studies acknowledge the difficulties adolescents experience in critically evaluating the information they find online, thus digital texts, images, audio, or videos presented on the Internet (Brand-Gruwel et al., 2005; Fraillon et al., 2015; Kanniainen et al., 2019; Kiili et al., 2008; McGrew et al., 2018; Mittermeyer, 2005; Pérez et al., 2018). Those studies provide ample evidence that young students, or ‘digital natives’ as some call them, do not possess the critical thinking skills they are assumed to have by nature of being frequently exposed to the Internet since they were little kids. In fact, CORE is a highly cognitive and complex skill that requires explicit and intensive instruction (Brand-Gruwel et al., 2005; Kanniainen et al., 2019; Leu et al., 2017). On the one hand, the CORE program seeks to increase the body of knowledge in new (digital) literacies by piloting an intervention to improve high school student's ability to evaluate online resources in the Quebec context. On the other hand, my study contributes to this overall goal by evaluating the program's implementation fidelity to gain knowledge about improving the delivery quality for subsequent implementations.

How to Evaluate the CORE Program

Program evaluation is fundamental to designing and implementing educational interventions. There are two distinct concepts embedded: program and evaluation. Let us separate these two components. First, *a program* is defined as a set of resources and activities directed to a particular goal, usually under the direction of a management team (Alkin & Vo, 2017; Hatry et al., 2015; Owen, 2006). In the evaluation terminology, the program is the *evaluand,* a generic name coined by Michael Scriven (Mertens, 2015). Second, *evaluation* can take diverse meanings depending on the context in which it is used, discipline, sector, and nation (Golden & Toledo Figueroa, 2020; Mertens, 2015). Some authors characterize evaluation as “distinguishing the worthwhile from the worthless, the precious from the useless” (Vedung, 2000, p. 12). A more straightforward definition, provided by Alkin & Vo (2017, p. 10), is

"considering the merit or worth of an entity," where the entity is the program subject to the evaluation. Moreover, there is a distinction between merit from worth (Mathison, 2005; Mertens, 2015). While *merit* is the absolute or relative quality of the evaluand, *worth* is the evaluand's value in a specific context.

A concept fundamental to program evaluation is that of stakeholders. A *stakeholder* is an individual who has a vested interest in the evaluation findings. These individuals can impact or be impacted by the evaluation process or its outcomes. Multiple stakeholders are involved in a program, including funders, administrators, facilitators, and participants. However, not all have the same level of participation and interest in the evaluation. *Primary or key stakeholders* are individuals with a strong interest in the program and the power to make or influence decisions. They are considered 'intended evaluation users' within the UFE framework[3]. *Other interested stakeholders* are individuals who may have concerns, interests, or different perspectives valuable to be considered during the evaluation process. Evaluators acknowledge the importance of working with the primary stakeholders to identify the purpose of the evaluation and the criteria for judging the program's success (Alkin & Vo, 2017; Bryson & Patton, 2015; Patton, 2008).

Making the Distinction Between Research and Program Evaluation.

Additional to the intrinsic definitions of program evaluation, this study situates program evaluation within the research context. Generally, both concepts denote a type of disciplined inquiry (Alkin & Vo, 2017); both are rigorous and *systematic* —planned and methodical— processes. According to Mathison (2008), there are four fundamental distinctions between research and program evaluation. One key difference is in the purpose. While evaluation seeks the determination of merit, worth, or value, research aims to understand the world. Second, decision-making is inherent in evaluation versus the "affirming or establishing a conclusion"

[3] This Utilization-Focused Evaluation (UFE) Framework is further explained.

(Mathison, 2008, p. 190) essential to research. Third, typically, research increases the body of knowledge in a specific discipline or topic, striving for generalization applicable across contexts, geography, and time (Alkin & Vo, 2017; Clark & Creswell, 2014; Mathison, 2008). In contrast, program evaluation looks for insights leading to action and decisions in a short- or long-term period: “this program, at this time” (Alkin & Vo, 2017, p. 9). Evaluation acknowledges that the same program may achieve different outcomes depending on the context of its application, with diverse participants and stakeholders involved in particular situations (Alkin & Vo, 2017; Hagerman, 2019; Mertens, 2015). Lastly, while evaluation includes traditional research methods, it has also developed its own methods to determine the value, merit, or worth. This study understands program evaluation as applied research, following a rigorous and systematic process and acknowledging the distinctions.

Program Theory and Logic Model Underlying the Implementation Evaluation

Bickman (1987) argues that every program has an implicit or explicit theory. Interventions are more frequently expected to specify a sound program theory to explicate the mechanisms through which a program will achieve its intended outcomes (Mowbray et al., 2003). A *program theory* explains how a program should work under certain conditions and the expected results (Bickman, 1987; Sidani & Sechrest, 1999). Some authors called this a theory of change; thus, what needs to be done, using given resources, by whom, to attain the anticipated outcomes in a specific context (McLaughlin & Jordan, 2015; Sidani & Sechrest, 1999). Depending on the evaluator's discipline and perspective, there could be more than one "correct" program theory, which implies that different constructs may account for the program effects. For example, the CORE intervention program has drawn its theory of change accounting for the quality of the intervention material, fidelity to conducting the intervention, engagement and motivation of the students, and preparation and confidence of the teachers. All these factors are hypothesized to contribute to the success of the program.

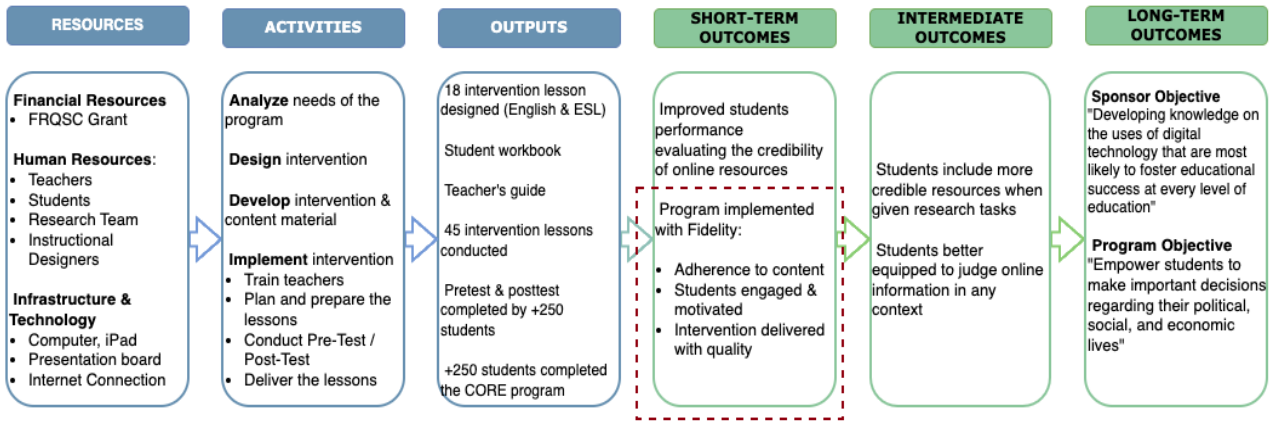
Accompanying the program theory are the *logic models*, one technique to make the

implicit program theory explicit (Bickman, 1987; Sidani & Sechrest, 1999). Logic models have been at the core of program evaluators for decades in various ways (McLaughlin & Jordan, 2015). The logic model establishes a framework for understanding the program's elements, the assumed relationships, and the role of context. According to McLaughlin & Jordan (2015), there are six components to describe the program: the resources, the activities, the outputs, and the short-term, intermediate, and long-term outcomes. The resources comprise human, financial, technology, and infrastructure or other inputs required to support the program. Activities are the essential action steps necessary to produce program outputs. The program outputs are the products, goods or services provided to customers or program participants. Outcomes represent 'evidence' that the activities were implemented as intended. Depending on the activity, the outcomes are intermediate deliverables for subsequent activities. Outcomes are the changes or benefits to the program's targeted people or organization.

This study is premised on a theory-based program evaluation, building upon the program theory and its logic model developed during the primary planning stage of the CORE program (See figure 1 Abbreviated Logic Model. See Appendix A for the full logic model version).

Figure 1

Abbreviated CORE Program Logic Model and Theory of Change



The following summarizes the underlying CORE program's theory of change:

> *"IF the research team designed and developed a high-quality CORE intervention material AND the teachers delivered the lessons with fidelity, THEN high school students will improve their ability to evaluate online information critically, and more importantly, will be empowered to make better-informed decisions in their lives" (*Jimenez et al., 2021, p. 4*)*

As the logic model illustrates, programs change and mature over time; thus, program evaluations can be conducted at different developmental stages (Alkin & Vo, 2017; CDC, 1999; Patton, 2008). Stakeholders might be interested in measuring the program's results in three phases: planning, implementation and effects. During the planning phase, program activities are untested, and an evaluation's purpose is to refine the plan. In contrast, activities have been taken to the field during the implementation and tested. An evaluation of the second phase is to compare ideal versus actual program activities and improve operations. This latter type of evaluation is known as *program implementation evaluation*. During the last stage, enough time has passed for the program's effects to emerge. This study's CORE program aims to conduct an implementation evaluation focusing on evaluating short-term outcomes (as opposed to intermediate or long-term outcomes.)

Utilization-Focused Evaluation (UFE) Guiding Evaluation Framework

Evaluations can be planned and conducted in many ways; however, the research team leaders directing the CORE program decided in early planning stages to adopt the Utilization-Focused Evaluation Framework (UFE) based on the principle that an evaluation should be judged on its usefulness to its intended users (Patton, 2008, p. 37). When designing a utilization-focused evaluation, the attention is focused on the intended use by intended users. Patton (2008) defines primary stakeholders as *intended users* of the evaluation. They have a direct and identifiable stake in the evaluation. They are encouraged to participate in the evaluation planning phase as an effective strategy to create ownership and increase the

probability that the evaluation results will be used (Paton, 2008; Ramirez & Brodhead, 2013). A utilization-focused evaluation is designed to answer the primary stakeholders' questions to decide the program's future, either continue, modify, or terminate, a.k.a. *intended use* (Hatry et al., 2015; Patton, 2008). The UFE framework does not advocate for particular content, method, or theory. It is a guiding framework rather than a methodology. It may include a variety of evaluation methods and paradigms. It is a situational and personal approach guided by an interactive process between the evaluator and the primary stakeholders. Following this principle, this study worked closely with the intended users (i.e., primary stakeholders), responding to their needs, situation, and context, increasing the likelihood of effectively using the findings and results. However, this does not mean the process is linear (Ramirez & Brodhead, 2013). UFE is summarized into a series of 12 steps (Figure 2) grouped into four phases: Preparation, Design, Analysis and Undertaking the Evaluation.

Figure 2

The UFE Framework in Steps

The CORE implementation plan developed by the research team during a previous phase included the execution from Step 1 (Assessing Program Readiness) until Step 9 (Collecting the Data). Step 10, Analyzing Data, entails the scope of this evaluation. Furthermore, Step 11 (Facilitation of Use) and Step 12 (Meta Evaluation) are out of the scope of the present study.

Program Evaluation Guided by Standards

The Joint Committee on Standards for Educational Evaluation (JCSEE) developed a set of 30 standards for practical evaluation, grouped into four categories: Utility, Feasibility, Propriety, and Accuracy (Yarbrough et al., 2010). A balance of these standards is recommended, thus "utility-focused, feasibility-conscious, propriety-oriented, and accuracy-based" (Patton, 1994, p. 195). Utility standards aim to increase the value for the stakeholder in using the evaluation results as they meet their needs. Feasibility standards are intended to increase evaluation effectiveness and efficiency (e.g., project management, practical procedures, resource use). Propriety standards underpin what is proper, fair, legal, correct and just in evaluation (e.g., responsive and inclusive orientation, human rights and respect, transparency and disclosure). Accuracy standards are intended to increase the dependability and truthfulness of the evaluation propositions and findings (e.g., justified conclusions and decisions, valid information, reliable information). Balancing these four standards is a crucial task during program evaluation. For example, an evaluation could be feasible and accurate but does not have the potential for use. In another case, the evaluation will serve the information needs of the stakeholders but is not feasible, or it might imply unethical issues. This study will be guided by these four standards, beginning with the premise that the findings will be of practical use to the stakeholders. Secondly, the evaluation has been deemed feasible by the primary stakeholders. Thirdly, it will be conducted legally and ethically. Lastly, the results will reveal and convey technically adequate information about the features determining the CORE program's worth.

Implementation Fidelity

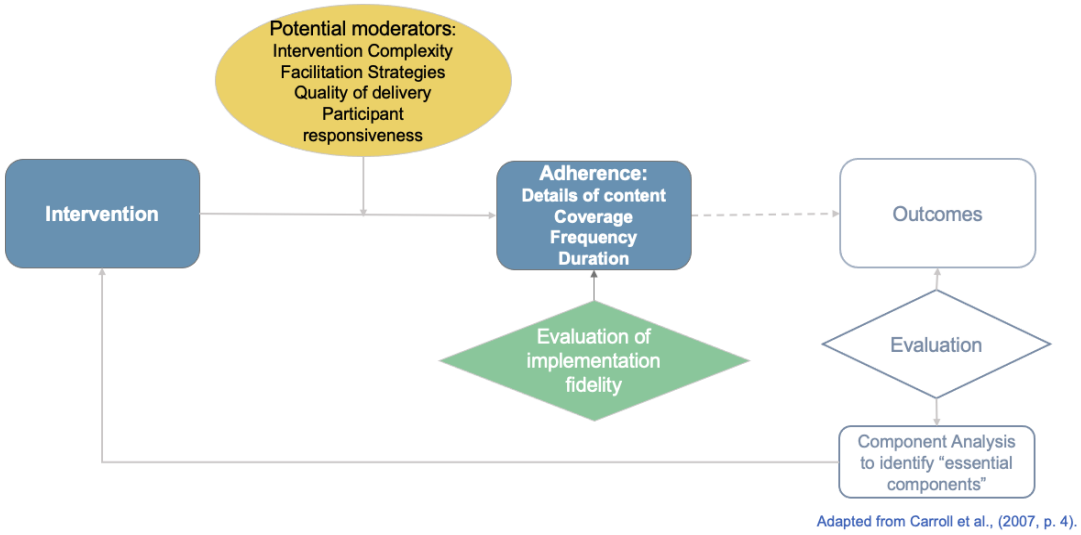
At the core of program implementation evaluation is the concept of *program fidelity*. Implementation fidelity is understood in this study as the degree to which an intervention program is delivered as expected by the program developer (Carroll et al., 2007; Mowbray et al., 2003; O'Donnell, 2008). Some authors refer to the same concept as treatment fidelity, treatment integrity, or fidelity of implementation (Breitenstein et al., 2010). In general, all these definitions are consistent. Evaluation of implementation fidelity is crucial because it may moderate the relationship between the intervention and the outcomes and helps understand the program's effectiveness (O'Donnell, 2008). It can give confidence to the researchers in attributing results and conclusions to the intervention and preventing potentially false conclusions about the program's effectiveness. Background literature on implementing with fidelity is not voluminous. It is primarily influenced by research in health science, where precise intervention delivery and accurate treatment description are critical for replication purposes (O'Donnell, 2008; Swanson et al., 2011). Research on program fidelity for interventions in digital literacy is scarce worldwide (O'Donnell, 2008). Implementation fidelity is a relatively new concept in K-12 intervention programs. However, it is not novel in the evaluation practice and is nowadays an expected piece of the quality of a program evaluation (O'Donnell, 2008; Mowbray et al., 2003). In a study conducted by Swanson et al. (2011), interested in examining fidelity reporting in intervention research for education, they concluded that 67% of articles published in high-impact general and special education journals (n = 76) collected and reported fidelity data for the interventions implemented. These results were encouraging since they were considerably higher than previously reported. The results Donnell (2008) obtained in her review study concluded that few studies correlated fidelity implementation with outcomes. However, the five studies that met the selection criteria "consistently showed statistically significantly higher outcomes when the program was implemented with greater fidelity" (Donnell, 2008, p. 50).

A review of the health literature indicates five criteria to measure implementation fidelity:

(1) adherence to content, (2) exposure, (3) quality of delivery, (4) participants' responsiveness, and (5) program differentiation (Carroll et al., 2007; O'Donnell, 2008). These criteria have been expanded and adapted to the educational domain, informing education intervention researchers' work (Swanson et al., 2011; Carroll et al.,2007.) The proposed framework for implementing fidelity, depicted in Figure 3, outlines the elements of implementation fidelity and their relationship. The dotted line represents that the relationship between the intervention and the outcomes is external to implementation fidelity. However, the degree of implementation fidelity attained can affect this relationship. Three basic elements are represented in the framework: adherence, potential moderators and essential components.

Figure 3

Conceptual Framework for Implementing Fidelity



Adapted from Carroll et al., (2007, p. 4).

Adherence is the bottom-line measurement of implementation fidelity. It is considered a mediating variable of the outcomes. Adherence includes the subcategories of content, frequency, duration and coverage (i.e., dose in the health domain). If the intervention adheres to the pre-determined details of the content, the coverage, the frequency, and the duration, then fidelity is likely high. The adherence level achieved may be affected or moderated by other variables: intervention complexity, facilitation strategies, quality of delivery and participant responsiveness. Achieving a high adherence and implementation fidelity level is difficult (Carroll

et al., 2007). Several factors can affect or moderate the results of an implemented intervention. In Carroll et al. (2007) framework, there are four basic moderators to the adherence variable: intervention complexity, facilitation strategies, quality of delivery, and participant responsiveness. The *intervention complexity* means how simple or complex the intervention is. It also includes how specific or vague it has been defined. When interventions have been detailed and specific, they have returned greater levels of implementation fidelity (Carroll et al., 2007) than vague ones. There is also evidence that simple interventions achieve higher levels of fidelity than complex ones. A complex intervention gives more room for modifications and deviations from the intended content and, as such, is more vulnerable to not being implemented correctly or completely. *Facilitation strategies* are defined in this conceptual framework as the support given to the people delivering the intervention in terms of training, monitoring and feedback to optimize and standardized the intervention as possible. It is possible, then, that these strategies moderate the fidelity level outcomes. The more is done in terms of training, support and feedback given to the intervention implementers, the higher levels of fidelity may be achieved. This factor is even more relevant for complex interventions, as simple interventions may not require major facilitation strategies. According to Carroll et al. (2007), some studies have shown the positive effect of specific facilitation strategies in implementing an intervention. However, “no study has yet measured the moderating effect of these strategies on the degree of implementation fidelity” (Carroll et al., 2007, p.6). *Quality of delivery* is another potential moderator of the relationship between the intervention and the implementation fidelity. It is defined as how the implementer delivers the program, thus using the prescribed techniques, processes, and methods. If the content of an intervention is badly delivered, then the implementation fidelity level might be negatively affected. *Participant responsiveness* refers to the extent to which participants are engaged and involved in the activities and content of the intervention. It includes participants’ judgement about the outcomes and relevance of the program. If participants perceived the intervention as irrelevant or unimportant, it might

negatively affect implementation fidelity. Moreover, participants in this context include those receiving the intervention and those responsible for it (e.g., teachers). Participants' beliefs and enthusiasm about the intervention may impact the implementation fidelity achieved. For example, the teachers' principles and opinions about drugs might affect a drug abuse prevention program in a school setting. The context, setting or organization may also influence the response of those delivering the intervention. If the organization leaders are not committed to the program, the responsiveness of the implementers might be affected. Lastly, component analysis refers to identifying the unique features of the program component that are essential for the intervention's success.

This study is guided by Carroll et al. (2007) proposed conceptual framework for implementing fidelity, delimiting its scope to the following components: Adherence to Content (mediator variable), Quality of Delivery (moderating variable), and Participant Responsiveness (moderating variable). The Adherence to Content variables includes Coverage, Frequency and Duration. Other moderating variables (e.g., facilitation strategies and intervention complexity) are excluded from the program's scope as well as the Program Differentiation element (i.e., identification of the 'Essential Components.')

Evaluation Questions

The purpose of this convergent parallel mixed method study is to evaluate the extent to which the CORE intervention was delivered with fidelity in the context of three high schools in Quebec. More specifically,

a. Did the teachers adhere to the content and duration of the intervention?

b. Did the teacher deliver the program with high quality? Specifically, how prepared, confident and enthusiastic were the teachers?

c. How engaged and motivated were the students? Are student and teacher participants satisfied?

d. Overall, what are the strengths and weaknesses of the program?

Chapter Three: Research Design and Methodology

This chapter details the research design used in this study. Then it describes the context of the analysis and an overview of the participants, data collection instruments, measures and procedures undertaken to respond to the research or evaluation questions.

A pragmatic convergent mixed methods approach (see Figure 4) was used to evaluate the CORE program's implementation fidelity. The pragmatism was inherited from the UFE framework that guided the entire CORE program evaluation, paying particular attention to providing primary stakeholders (i.e., research team leads) helpful information to empower their decision-making regarding the CORE program's future (Patton, 2008). Quantitative and qualitative data were concurrently collected for five interventions conducted by four teachers in three secondary schools in Quebec, Canada. Furthermore, data were collected from multiple sources, capturing the voice of teachers, students, and researchers. Triangulating quantitative and qualitative data facilitated the validation through cross-verification from multiple sources (Figure 5). The value of this design is that a more comprehensive understanding and explanation of the program results were obtained through multiple methods and sources (Miles & Huberman, 1994).

Figure 4

The Convergent Parallel Mixed Methods Design (Clark & Creswell, 2014).

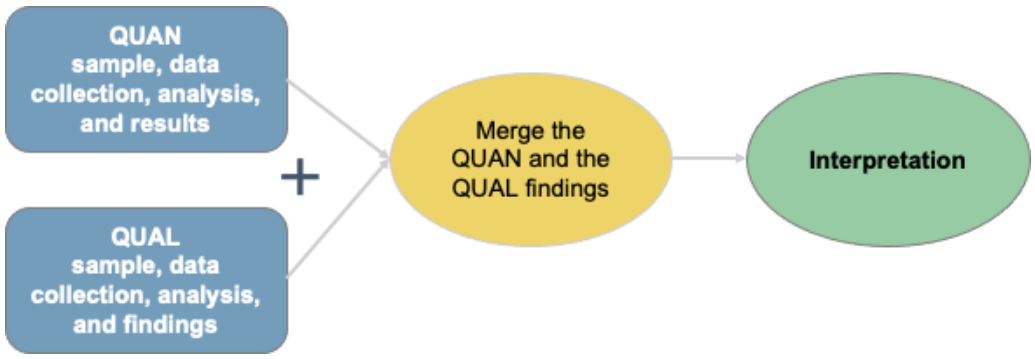
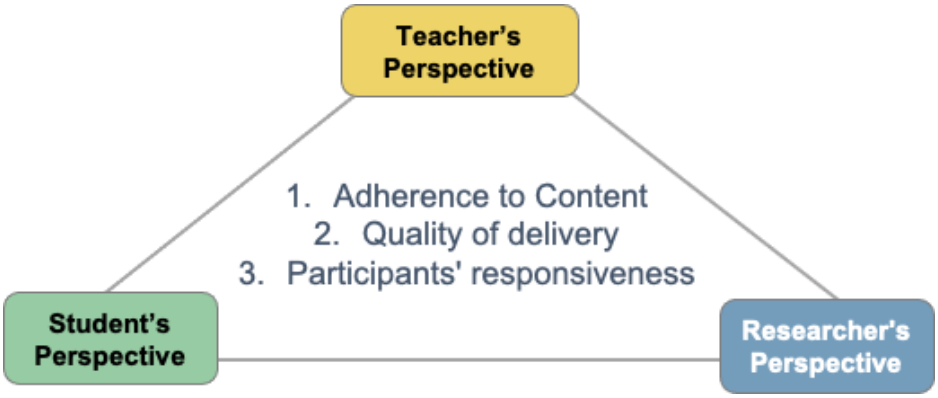


Figure 5

Data Triangulation – Visual Representation



Quantitative analysis included descriptive statistics of trends and a one-way ANOVA test to determine whether student engagement and motivation scores differed for classrooms with different teachers. The qualitative analysis used thematic analysis resulting in a detailed exploration of multiple cases. Five cases were defined bounded by the classroom where the CORE intervention was conducted. One teacher taught the lessons in two different classrooms, thus making two distinct cases for the same educator. Regarding the time boundary, all teachers completed the nine lessons intervention at different times during the first semester of 2022. Some teachers conducted the intervention during winter, while others during spring. A cross-case analysis was conducted to compare the five cases and find similarities and differences.

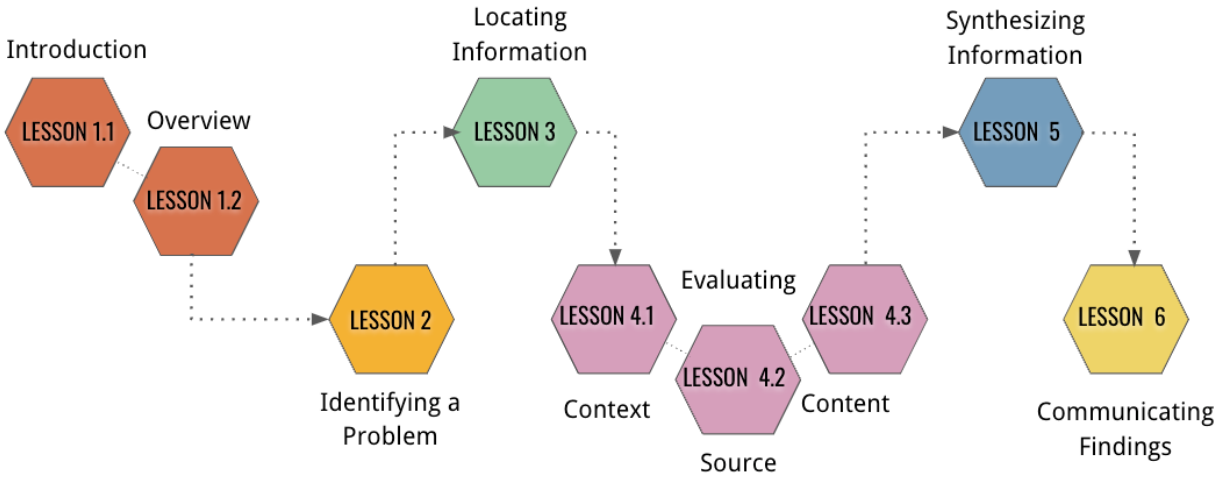
The combined quantitative and qualitative analysis included the three constructs in the adopted Carroll's Fidelity Framework (i.e., adherence to content, quality of delivery, and participants' responsiveness) plus new emerging themes (e.g., content adaptation, classroom context, intervention complexity). Criteria were established to determine if the intervention was delivered with an acceptable level of fidelity based on previously reported levels of fidelity in the literature (An et al., 2020; Toomey et al., 2017), followed by a validation process with the primary stakeholders. All the quantitative scores were categorized into three levels according to

the predefined criteria: 80% to 100% was interpreted as high level, 51% to 79% as moderate level, and 0% to 50% as low level.

Six modules and nine lessons encompassed the intervention (Figure 6). Two distinct CORE intervention versions were developed and implemented according to the school's language of instruction. The English (ENG) version was implemented in two anglophone schools where English was the language of instruction. The English as a Second Language (ESL) version was used in one francophone school where English was taught as a second language. The intervention content for the ESL was adapted to include notes from the Progression of Learning at the Secondary Level English as a Second Language Core Programs (Ministère de l'Éducation et Ministère de l'Enseignement supérieur, n.d.) Each lesson lasted 55 minutes for the schools with English as a language of instruction and 75 minutes for English as a Second Language (ESL) schools.

Figure 6

Intervention Program Content Structure: Six Modules, Nine Lessons



Participants

The primary stakeholders —intended users in the evaluation jargon— were the research team leaders, which was comprised of three researchers assigned to the Department of Education of the University of Concordia and a practicing teacher in a private school in

Montreal, Canada. The principal investigator (PI) was Dr. Julie Corrigan, Assistant Professor of Digital Literacies at the University of Concordia. Two co-investigators worked in partnership with the PI in the management, development, and execution of the project: Dr. Kim McDonough, Professor and Canada Research Chair in Applied Linguistics at Concordia, and Dr. Heike Neumann, Senior Lecturer in English as a Second Language at Concordia. In addition, a teacher practitioner, Andrea Barrios, Secondary English and Spanish teacher at a francophone private school, contributed her professional experience gained in practice and teaching to design and deliver the intervention. These lead participants were strongly interested in the CORE program and had the power to make or influence decisions about its future, as informed by this evaluation results.

Four teachers conducted the intervention program in three high schools (Table 1). One teacher in a francophone private school implemented the ESL version in two different classes. One anglophone public school implemented the ENG version in two different classes by two different teachers. One anglophone private school implemented the ENG version. For the purposes of confidentiality, each participant was given a pseudonym, and school names and exact locations were not disclosed.

Table 1

Participating Schools and Classes

Teacher	School type	Language of instruction	Grade	Class	Class subject	Number of participating students
Kasey	Public	English	Grades 10 (Sec IV)	N/A	ELA	116
Malena	Public	English	Grade 11 (Sec V)	N/A	ELA	59
Amy (51)	Private	French	Grade 11 (Sec V)	51	ESL	32
Amy	Private	French	Grade 11 (Sec V)	55	ESL	36
Chris	Private	English	Grade 9 (Sec III)	N/A	History	20

Note. ESL = English as a Second Language; ELA = English Language Arts.

A total of 263 (N = 263) students from Grade 9 (Sec III), Grades 10 (Sec IV) and Grade 11 (Sec V) participated in the CORE intervention program and were observed during the classes. Parental consent and student assent were obtained from 128 students. Of these students, n = 95 (74.2%) completed the student engagement survey, making this our final sample size. Our sample represents a broad diversity (ethnically, linguistically, and by grade and gender) of students (Table 2).

Table 2

Demographic Characteristics of Participating Students

Baseline characteristic	n = 95	Percentage %
Gender		
Female	48	50.5
Male	41	43.2
Unknown	6	6.3
Language		
English	44	46.3
English and/or French and other	14	14.7
Other	13	13.7
Both English and French	10	10.5
French	8	8.4
Unknown	6	6.3
Racial Identity		
White	58	61.1
Self-Identified as a Person of Colour	19	20.0
Don’t know/Prefer not to say/Unknown	18	18.9
School type		
Public	76	80.0
Private	19	20.0
Grade		
SEC 5 / 11th grade	58	61.1
SEC 4 / 10th grade	31	32.6
SEC 3 / 9th grade	6	6.3

Five research assistants (RA) from the Department of Education at Concordia University accompanied and observed the four teachers throughout the delivery of the entire intervention (Table 3). Their role was relevant for this study since they represented the research perspective by attending the interventions in the classroom settings and recording their observation notes about the program fidelity; thus, they recorded scores and notes for adherence to content, quality of delivery, and participants' responsiveness. It is important to note that research assistants were assigned to the observation task according to availability, so none of the teachers had the same observer during the entire intervention.

Table 3

Research Assistants (Observers)

Teacher	Lesson 1.1	Lesson 1.2	Lesson 2	Lesson 3	Lesson 4.1	Lesson 4.2	Lesson 4.3	Lesson 5	Lesson 6
Kasey	Ellen	Talya	Talya	Tristen	Talya	Tristen	Talya	Talya	Tristen
Malena	Talya	Ellen	Talya	Ellen	Talya	Talya	Talya	Ellen	Ellen
Amy (51)	Aria	Aria	Kelly	Kelly	Aria	Aria	Aria	Aria	Aria
Amy (55)	Aria	Aria	Kelly	Kelly	Aria	Aria	Aria	Aria	Aria
Chris	Tristen	Tristen	Tristen	Ellen	Tristen	Tristen	Tristen	Ellen	Ellen

Data Collection Instruments

Three data collection instruments were used to gather evidence: Classroom Observation Protocol, Teacher's Journal Log, and Student's Engagement and Motivation Survey. These instruments captured quantitative and qualitative data from multiple perspectives (i.e., teachers, students, and researchers). A summary of the evaluation questions, data sources, participants and methods is shown in Table 4.

Table 4

Evaluation Questions, Participants, and Methods

Evaluation questions	Data Sources	Participants and methods
Did the teachers adhere to the content and duration of the intervention?	● Observation Protocol ● Teacher's journal log (Questions 5, 6 & 7)	● Classroom observations with each participating teacher: 45 (nine lessons x five interventions) ○ Each RA logged their ratings and perceptions about how teachers adhered to the lesson content. ● Teacher's journal log: 45 (nine lessons x five interventions) ○ Each teacher logged their ratings and experiences with program adherence.
Did the teacher deliver the program with high quality? Specifically, how prepared, confident and enthusiastic were the teachers?	● Observation Protocol ● Teacher's journal log (Questions 8, 9, 10 & 11)	● Classroom observations with each participating teacher: 45 (nine lessons x five interventions) ○ Each RA logged their ratings and perceptions about teachers' preparation, confidence, and enthusiasm to teach the lesson. ● Teacher's journal log: 45 (nine lessons x five interventions) ○ Each teacher logged their ratings and perceptions about their preparation and confidence.

Evaluation questions	Data Sources	Participants and methods
How engaged and motivated were the students? Are student and teacher participants satisfied?	● Observation Protocol ● Teacher's Journal Log (Questions 12 & 13) ● Student's Engagement and Motivation Survey	● Student's Engagement and Motivation Survey: 95 students (with parental consent and assent) ○ Students completed a self-report Engagement and Motivation Survey during the last lesson. ● Teacher's Journal Log: 45 (nine lessons x five interventions) ○ Each teacher logged their ratings and perceptions about students' engagement and motivation. ● Classroom observations with each participating teacher: 45 (nine lessons x five interventions) ● Each RA logged students' engagement and motivation, including ratings and perceptions during the lesson taught.
Overall, what are the strengths and weaknesses of the program?	● All data sources	● Consolidated qualitative and quantitative analysis considering all perspectives (i.e., RAs, teachers and students) by using a data triangulation strategy

Note. RA = Research Assistant

Classroom Observation Protocol

The classroom Observation Protocol served the dual purpose of collecting qualitative and quantitative data about the teacher's adherence to the content, how well the teacher delivered the lesson and how motivated and engaged the students were. The research

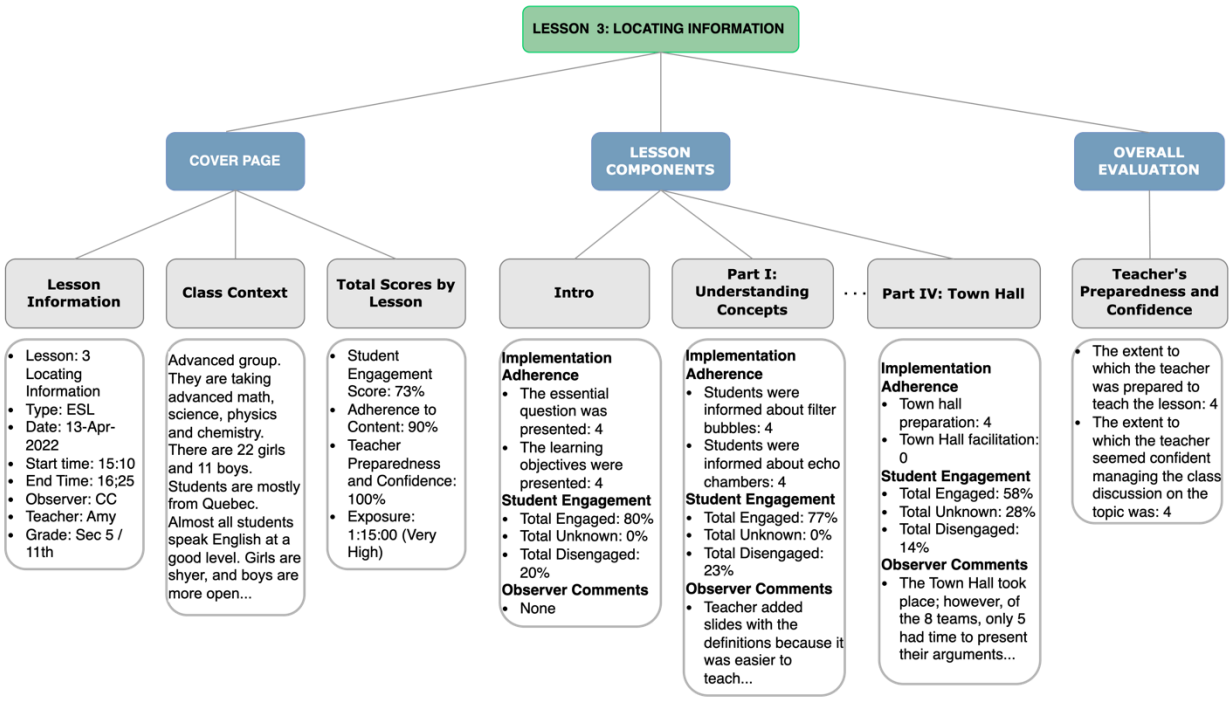
assistants completed one instance of the Observation Protocol for each observed class. The quantitative part of the protocol that captures student engagement and motivation was built upon the Behavioral Engagement Related to Instruction (BERI) protocol developed by Lane & Harris (2015). This protocol aims to target a sample of approximately ten students. The observer quickly (in two-minute intervals) and cyclically writes down whether said students were engaged, disengaged, or unknown behaviour (when the view was unclear) in succession. The level of engagement could be used to determine the intervention activities that most effectively engage students and provide valuable feedback to the stakeholders. The BERI protocol has been demonstrated to be reliable for collecting classroom observation engagement behaviours: “Data from 2,154 judgments of individual student engagement, from six pairs of observers in three different educational settings, with five different instructors were used to evaluate interrater reliability. The average interrater agreement was calculated to be 96.5%” (Lane & Harris, 2015, p. 87).

To give the RAs a means to easily record the data, the Observation Protocol tool was developed in Microsoft Excel. Its general structure entailed three segments: (1) cover page; (2) lesson parts; and (3) overall lesson evaluation. Figure 7 depicts an example of the Observation Protocol for *Lesson 3 – Locating Information* in one of the classroom interventions.

Cover Page. The cover page captured general information about each lesson, such as lesson number and title, program type (ENG/ESL), class start and end time, teacher’s name, and observer’s name. It also included the classroom context field where the observer (i.e., research assistant) described the specificities of that classroom (e.g., number of boys and girls, relationship with the teacher, general environment, technology used, etc.). A consolidation score table from the lesson was automatically generated by averaging the scores of the parts and components of the lesson. The last aspect of the cover page was a field for the observer to record other comments relevant to the research. For example, any deviation from the original content, absenteeism, or any additional information deemed pertinent by the observers.

Figure 7

Observation Protocol Example for Lesson 3 – Locating Information



Lesson Parts and Components. Each lesson was divided into several parts. The number of parts was variable depending on the lesson. For example, *Lesson 2 – Identifying a Problem* was comprised of five parts: (1) Intro, (2) Self-Assess Knowledge and Beliefs, (3) Understanding Terminology, (4) Town Hall Preparation, and (5) Reflecting on Confirmation of Bias. Each part was further divided into smaller components or instructional activities to be delivered by the teacher (e.g., present the learning objectives, explain terminology, conduct activity). Each instructional activity became the minimum unit of quantitative measure for Adherence to Content. The observer recorded a quantitative rating for student engagement in each lesson part[4]. Furthermore, the observer had the opportunity to note any significant teacher or student comments regarding the implementation, observations about how the teacher delivered the content and how the student responded to the intervention. For example, a

[4] The scoring and calculations for adherence and student engagement are further explained in the Measurements section.

teacher skipping an instructional activity, adding extra explanations or examples, students disengaged by any specific circumstance, or observers unable to capture information. The information recorded in this field became the input for the qualitative analysis.

Overall Lesson Evaluation. The observer used the last tab in the Observation Protocol to capture quantitative ratings about teachers' preparedness, confidence**,** and enthusiasm, as well as final qualitative comments about the lesson's quality of delivery.

Teacher's Journal Log

A journal or log is a structured diary instrument for collecting data continuously (Alkin & Vo, 2017). The teacher's Journal Log was a self-reported instrument developed on the Google Form platform. By completing this journal log, teachers supplied the research team with ratings (quantitative data) and comments (qualitative) about their perception of adherence to content quality of delivery and student responsiveness. The teacher completed one form after each lesson was taught. At the end of the journal, teachers could record additional comments about the lesson, for example, if the teacher did not have time to complete the lesson and the reasons why, any changes or deviations from original content and the rationale for those modifications. These notes became data for qualitative analysis.

Student's Engagement and Motivation Survey

The Student's Engagement and Motivation Survey captured students' self-reported ratings, perceptions, and experiences about the program's relevance to them and their motivation and engagement with the program. This survey was anonymous and adapted from the Academic Engagement Scale of the Consortium on Chicago School Research Biennial Survey (CCSR/AES) developed by the CCSR (http://ccsr.uchicago.edu) (Fredricks et al., 2011). The survey developers reported that this instrument complies with the validity and reliability standards with a reported Cronbach's alpha of .65 to .68. The survey integrated four open-ended questions to capture details about what the students like the most and the least of the CORE program. In addition, the research team wanted to explore students' topics of interest to

be considered for the next intervention iteration and other general comments students wanted to share.

Measurements and Themes

Based upon the conceptual framework for implementing fidelity proposed by Carrol et al. (2007), this study measured three components: adherence to content, quality of delivery and participant responsiveness. Adherence to content (including frequency and duration) is essentially the bottom-line measurement of implementation fidelity, “the degree to which the intended content of an intervention is implemented is the degree of implementation fidelity achieved for that intervention” (Carrol et al., 2007, p. 3). We operationalize adherence to the content by capturing how close the teachers delivering the intervention adhered to the program as the CORE research team outlined. If the intervention adheres to the predefined details of the content, the frequency, and the duration, then fidelity can be said to be high. For the CORE program, the frequency was defined by nine lessons, with a duration between 55 to 75 minutes each (depending on the language of instruction, 75 minutes for ESL and 55 minutes for ELA). All teachers involved in the CORE intervention delivered all intervention lessons during the estimated time allocation; thus, the frequency and duration indicators were recorded as 100%. For operationalization and practical purposes, only the adherence to content scores (i.e., excluding frequency and duration) was used for the data analysis, consolidation, interpretation, and reporting of the results on this variable. The other two moderator variables we measured, according to the adopted framework, were quality of delivery and participant responsiveness. The quality of delivery was operationalized by measuring teachers’ preparedness, confidence, and enthusiasm to deliver the intervention content. In addition, participant responsiveness was operationalized as how engaged and motivated the students were during the intervention.

All the data for adherence to content, quality of delivery and participant’ was captured from three different sources or perspectives: teacher, researcher (i.e., research observer), and student. The instruments used for collecting the data differed for each source. Thus, teachers

completed the Journal Log, researchers used the Observation Protocol, and students completed the Engagement and Motivation Survey. The following sections detail each measurement scoring, calculation, and data source.

Adherence to Content

Teacher Perspective. The adherence to content measured by the teacher was recorded in the teacher's journal. A 5-point Likert scale (1 = s*trongly Disagree* to 5 = *strongly agree*) was used to capture the scores. The total points awarded through this process were divided by the possible points to produce a percentage score for each lesson. Classroom level adherence score was computed by averaging adherence scores of the nine lessons.

Researcher Perspective. The adherence to content measured by the observer or research assistant was recorded in the Observation Protocol. A simple scoring system was used. Each instructional activity within individual parts was allotted a maximum of 4 possible points. Only a score of 0 points was awarded if an entire activity was omitted. From 1 to 4, points were awarded based on how well the teacher adhered to the content designed (1= *Poor*; 2 = *Adequate*, 3 = *Good*; 4 = *Excellent*). The total points awarded through this process were divided by the possible points to produce an adherence percentage score for each part. Adherence score by lesson was computed by averaging all lesson parts' adherence scores. Classroom level adherence score was computed by averaging adherence scores of the nine lessons.

Quality of Delivery

Teacher Perspective. The quality of delivery measured by the teacher was recorded in the Teacher's Journal. A 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*) was used to capture the scores. The total points awarded through this process were divided by the total possible points to produce a quality of delivery percentage score for each lesson. Classroom level quality score was computed by averaging quality scores of the nine lessons.

Researcher Perspective. The quality of delivery measured by the observer or research

assistant was recorded in the observation protocol. These measurements included preparedness to teach the content, confidence, and enthusiasm that the teacher demonstrated during each lesson taught. A simple scoring system was used. From 1 to 4, points were awarded based on how well prepared, confident, and enthusiastic the teacher was (1= *Poor*; 2 = *Adequate*, 3 = *Good*; 4 = *Excellent*). The total points awarded through this process were divided by the possible points to produce a percentage score for each lesson. Classroom level quality of delivery score was computed by averaging quality scores of the nine lessons.

Participant Responsiveness

Teacher Perspective. The participant responsiveness measured by the teacher was recorded in the teacher's journal. A 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*) was used to capture the score of engagement and motivation. The total points awarded through this process were divided by the possible points to produce a participant responsiveness percentage score for each lesson. Classroom-level participant responsiveness score was computed by averaging the motivation and engagement scores of the nine lessons.

Researcher Perspective. Participant responsiveness was calculated in this study by measuring students' engagement and motivation based on the BERI protocol. The research assistant (i.e., observer) selected, at the beginning of the class, ten students to be observed, then she recorded the demonstrated behaviours in the Observation Protocol during each lesson part and instructional activity. Each student was rated as engaged, disengaged, or unknown behaviour (when the view was unclear) with the specific activity. A score of 1 point was awarded only the student was engaged and 0 points for disengaged or unknown. The total points awarded through this process were divided by the total possible points —excluding the unknown score (i.e., the observer could not record a value)— to produce a percentage score for each part. Engagement and motivation score by lesson was computed by averaging scores of all lesson's parts. The nine lessons' adherence scores were computed by averaging classroom-level engagement and motivation scores.

Student Perspective. Students completed the Engagement and Motivation Survey which encompasses 14 questions. A 5-point Likert scale (1= *strongly disagree* to 5 = *strongly agree*) was used to capture the scores. Three items in the scale that were negatively worded were reversed. Scores from the items were added to give each student an overall score. A percentage calculation was done by dividing the total score by the total possible points. The classroom-level participant responsiveness score was computed by averaging the motivation and engagement scores of all students in the classroom. The internal consistency of the adapted scale was checked by running a reliability of scale test. The scale had a high level of internal consistency, as determined by a Cronbach's alpha of 0.823.

Chapter Four: Findings

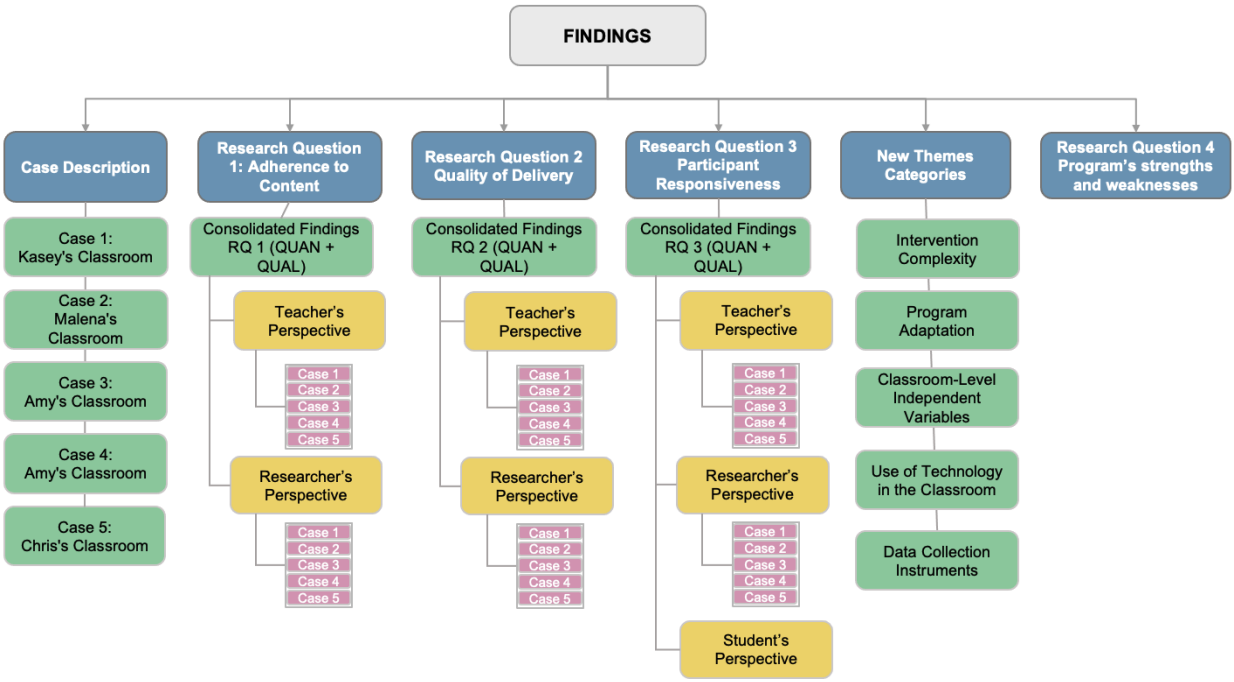
Following the pragmatic mixed method design that included the concurrent collection of quantitative and qualitative data, this study integrated the quantitative and qualitative findings to develop a deeper understanding of the CORE program fidelity. This chapter presents the results based on the data analyses described in the previous chapter *Research Design and Methodology*. Figure 8 illustrates the structure for displaying the findings. First, a description of the five cases is provided, relating the key characteristics of each classroom setting. Then, the results are displayed by each research question. Under each research question, a consolidated view of the findings is presented from three perspectives (i.e., teachers, researchers, and students). Next, detailed findings by perspective are presented using *joint displays,* a figure representing the quantitative (line graphs) and qualitative (illustrative quotes) findings side by side (Fetters, 2020). This integrated display format provides a more comprehensive and nuanced analysis and facilitates the researchers and readers to interpret and comprehend the findings. Lastly, we reported the new themes that emerged from the qualitative analysis and the program's strengths and weaknesses.

The quantitative scoring recorded in the three collection instruments (i.e., Observation Protocol, Teacher's Journal and Student Engagement and Motivation Survey) were transferred, analyzed, and consolidated in Microsoft Excel. In addition, IBM SPSS Statistics 29.0.0.0 was used to run the statistical tests (i.e., scale reliability, ANOVA). All notes recorded in the three collection instruments were transferred and consolidated into Microsoft Word and Excel to produce the qualitative data analysis. During the first analysis round, emerging themes were identified. A colour-coding scheme was used to highlight the emerging themes in the text. Comments (i.e., memos) were used throughout the coding process to document the processes and the data perceptions. A second round of analysis was done for verification and refinement. Finally, the free Worditout application was used to generate illustrative word clouds, presenting the students' responses to the open-ended questions (i.e., what they liked the most about the

CORE program, what they liked the least and topic recommendations)

Figure 8

CORE Findings Reporting Structure



Description of the Cases

Case 1 – Kasey's Classroom

Kasey is an English language arts professor for grade 10 (Sec IV) in an anglophone public school in Montreal. At this school students can enroll in three French programs, Regular (français langue seconde), français immersion and français langue d'enseignement. Kasey conducted the intervention in February. One hundred sixteen students participated in the CORE intervention; however, only 53 signed consent and assent forms. There was no fixed class schedule during the intervention. Some classes were in the morning, and others in the afternoon. The teacher provided laptops to students; if not enough, they used phones. The teacher made the material available on Google Classroom. During the first lesson, the observer (i.e., research assistant) noted a good rapport between Kasey and her students: "it was obvious that there was a positive relationship between teacher and students" (Ellen). Kasey linked the

intervention to a persuasive paper that the class was doing. She used this assignment instead of the “Independent Project” activity suggested in the CORE intervention. Students were allowed to solidify their topic and start their research after the pretest.

Case 2 – Malena’s Classroom

Malena is an English language arts professor for grade 10 (Sec IV) in an anglophone public school in Montreal. At this school students can enroll in three French programs, Regular (français langue seconde), français immersion and français langue d'enseignement. Malena conducted the intervention in February. There was no fixed class schedule during the intervention. Some classes were in the morning, and others in the afternoon. Fifty-nine students participated in the CORE program, with 41 signing the consent and assent forms.

Case 3 – Amy’s Classrooms 51

Amy is an English and Spanish teacher in a private francophone school in an off-island suburb of Montreal, southwestern Quebec. The school offers five programs: Performing Arts, World and Environment, Science, Sports and Educational Support. All students and staff are provided with iPad tablets, integrating technology into the service of teaching and learning. Amy conducted the CORE intervention in four English as a Second Language (ESL) classes for grade 11 (Sec V), but the research team observed only two of them: Group 51 and 55. The intervention was conducted in springtime from the beginning of April to the beginning of May. Some classes were taught in the morning and others in the afternoon. There was no fixed schedule. The Teacher allowed the students to bring and use their mobile devices during class. The Teacher used the Showbie app for the class, an easy-to-use classroom management platform that helps teachers bring their classrooms together. It combined all the essential tools for assessments, feedback, and communication. The Teacher uploaded the content and assignments into Showbie in advance, facilitating intervention. The students spoke primarily French when conversing with each other. The Teacher frequently reminded them they should speak English in her class, and they obeyed momentarily. Sometimes, the Teacher translated

concepts into French words when she noticed the students were not grasping the main concept. Likewise, the students who participated introduced words in French because they could not find the appropriate word in English.

Group 51 was an advanced class. They were taking advanced math, science, physics, and chemistry. Thirty-two students participated in the CORE program. From the total participating students, 14 signed consent and assent forms. Students were mostly from Quebec; thus, there was not much diversity. Almost all students spoke English at a good level. Girls were shyer, and boys were more open. This group was more disciplined than group 55, and students were quieter and more engaged.

Case 4 – Amy's Classrooms 55

Group 55 was a regular class (i.e., not advanced). Thirty-six students attended the CORE program. Only four students signed the consent and assent forms. Students had diverse backgrounds and mother tongues, unlike Group 51, where most were from Quebec. Almost all students spoke English at a good level. Girls were more open than boys. This group has taken Communication and Media classes and learned some CORE concepts.

Case 5 – Chris' Classroom

Chris is a history teacher for grade 9 (Sec III) in a private anglophone school in Montreal. The school offers co-educational programs from kindergarten to grade 12 (university preparatory) and extensive co-curricular offerings, including athletics, arts, service, debating, and leadership. It also offers an International Baccalaureate program. The intervention was conducted from mid-April to mid-May. The teacher expected some absences due to the Passover holiday celebration. There was no fixed class schedule during the intervention. Some classes were in the morning, and others in the afternoon. A total of 20 students attended the class. However, 13 signed the consent and assent forms.

Research Question 1: Adherence to Content

We defined adherence to the content in this study as the variable that determines the levels of the implementation fidelity of the intervention, i.e., how close the teachers delivering the program adhered to the program as delineated by the CORE team. Adherence to content is a mediating variable of implementation fidelity. If the intervention adheres to the predetermined specifics of the content, the frequency, and the duration, then fidelity can be interpreted as high. The frequency was defined by nine lessons, delivered according to the school's regular class schedule, and the duration was 75 minutes for ESL and 55 minutes for ELA. All teachers involved in the CORE intervention delivered all lesson of the intervention during the estimated time allocation. Thus, both the frequency and duration indicators were recorded as 100%. For operationalization purposes, only adherence to content scores were used for the data analysis to consolidate, interpret, and report the results on this variable. Note that we initially excluded two moderating variables from operationalizing adherence to content: intervention complexity and facilitation strategies. However, intervention complexity emerged as a theme during the qualitative analysis.

Overall, teachers had a high adherence to content level from a consolidated perspective with a mean of 79.9% from the teacher’s perspective, and 89.0% from the researcher’s view, as shown in Table 5. The strongest adherence to content level was found for Amy (51) (91.4%). Three moderating factors might explain this latter score. Firstly, both the researcher and Amy (51) reported minimal changes to the content of the intervention. Secondly, Amy (51) delivered the intervention for the second time, learning from the first time and accommodating the flow of the lessons, the material, and the timing for the subsequent class. Thirdly, Amy (51) teaches an ESL class with more class time (75 minutes) to deliver the content than the other teachers in the anglophone setting (55 minutes). Conversely, Malena achieved the lowest adherence to the content score of 75%, mainly driven by her self-assessment (67.8%) since the level was high from the researcher's perspective (82.1%). Three factors might explain Malena's score. Firstly,

she made a significant adaptation by replacing the Independent Project with the 'Persuasive Paper' project students had been working on in her English class, so she skipped activities related to the Independent Project. Secondly, she blended *Lesson 5 – Synthesizing Information* and *Lesson 6 – Communicating Findings* into one lesson, lacking time to deliver the entire material for these two lessons, aggravated by a disruptive class environment that specific day. Thirdly, time limitation might account for the lower adherence score, noted by both the researcher and the teacher.

Table 5

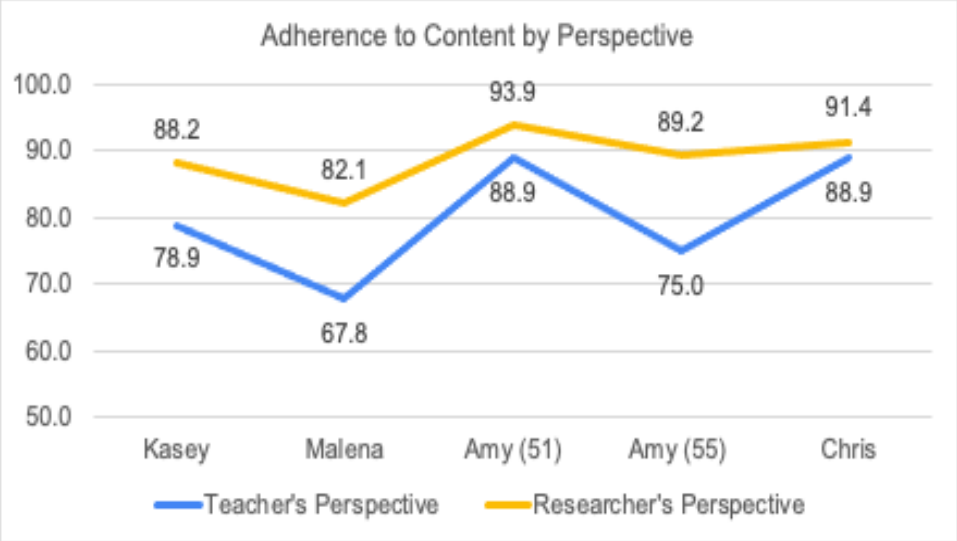
Consolidated Teacher's Adherence to Content Scores

Teacher	Consolidated Adherence to Content Score (%)		
	Teacher's Perspective	Researcher's Perspective	Difference between Perspectives
Kasey	78.89	88.22	9.33
Malena	67.78	82.11	14.33
Amy (51)	88.89	93.89	5.00
Amy (55)	75.00	89.22	14.22
Chris	88.89	91.44	2.55
M (Perspective)	79.89	88.98	9.09
Total Adherence Level by Perspective	Moderate	High	

Note. Adherence levels: High 80% – 100%; Moderate 51% – 79%; Low 0% – 50% (An et al., 2020; Toomey et al., 2017.)

Figure 9

Adherence to Content by Perspective



Comparatively, the mean of the researcher's perspective regarding adherence to content was higher (89%) than the teachers' perspective (79.9%) (See Figure 9). Malena's case reported the highest difference between the researcher's and teacher's perspectives (14.3%) in the content adherence scores, followed by Amy (55) with a difference of 14.2%. On the contrary, Amy's (51) case reported a slight difference (5%) between these two perspectives. These differences could probably be explained by the characteristics of the two instruments used to measure the construct. On the one hand, the Observation Protocol had more granular data collection points, thus at the activity and part level, while the Teachers Journal collected data at the lesson level. Collecting data at a more granular level might lead to better results. On the other hand, the Teacher's Journal was a self-assessment instrument, while the Observation Protocol was an external assessment. A plausible explanation for why self-assessment results were lower than external research observer results might be the subject of future research.

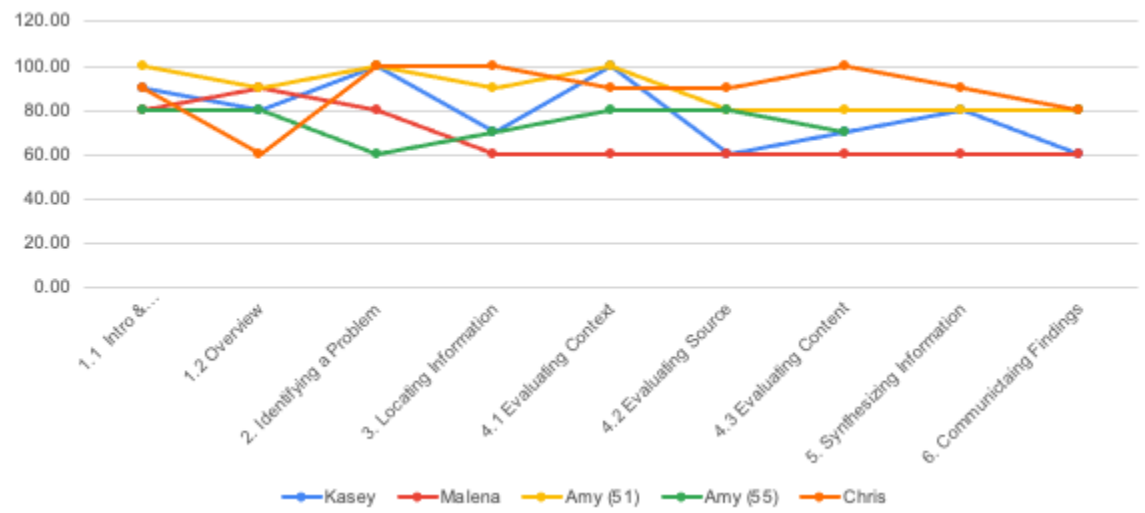
Adherence to Content – Teacher's Perspective

Teachers perceived a moderate adherence to content level (79.9%) from their consolidated perspective (Figure 10, Table 6). All teachers reported adaptations to some extent to the instructional content according to the grade level, the prior knowledge, time constraints or the contextual needs of the students. Chris and Amy (51) self-reported the highest levels of adherence to content (88.9%) throughout all the lessons. On the contrary, Malena was the teacher who self-reported the lowest adherence to content score (67.8%), recording 60% of adherence in six of the nine lessons. The teachers reported *Lesson 1.1– Intro & Pretest* and *Lesson 2 – Identifying a Problem* as the lessons they followed closely the most, scoring 88%. Conversely, they reported *Lesson 6 – Communicating Findings*, as the lesson they adapted the most with an adherence score of 72%, most likely explained by the fact that all teachers previously taught this topic. As an anecdotal highlight, but important for future implementation considerations, Chris had to improvise *Lesson 1.2 – Overview* as a reaction to the lack of Internet and WIFI at the school that day, but still perceived a moderate level of

adherence (60%), which might demonstrate his preparation, confidence and facilitation skills.

Figure 10

Teacher Adherence to Content Scores – Teacher's Perspective



Illustrative Quotes:

"I decided to follow the workbook rather than the slides and initiated the Town hall activity before teaching the concepts [*Lesson 3 – Locating Information*], allowing me to expand on how they found their sources. I thought it was a better approach" (Kasey).

"The logistics of the class [*Lesson 4.2 – Evaluating Source*] were affected as I did not feel as well prepared" (Kasey).

"The lesson examples should be more geared to what we are doing in class. Too much information about Covid. It would be more useful to have a way to incorporate what the kids are researching for their speech" (Malena).

"The lessons also could be combined [*Lesson 5 – Synthesizing Information* and *Lesson 6 – Communicating Findings*]" (Malena).

"I had more time to go over the intonation focus, I provided other examples comparing to French intonation of French cognates, and I explained the difference depending on the type of words (e.g., a record vs to record, a photograph vs a photographer)" (Amy [51]).

"The lesson [*Lesson 6 – Communicating Findings*] went well because explaining a thesis: claim and two reasons was a reinvestment of what we have worked at all year" (Amy [51]).

"The lesson went well, but we started a little late. Then it took a long time to examine explain their results about school closures, and then explain the key concepts: perspective, beliefs and confirmation bias. All of this took longer than expected because I elicited or provided examples and many students were engaged and participated" (Amy [55]).

"Perhaps have the town hall first before going into the rest of the lesson so all students have time to speak without being rushed [*Lesson 3 – Locating Information*]" (Chris).

Table 6

Teacher Adherence to Content Scores – Teacher's Perspective

Teacher	Total Adherence to Content Score by Lesson (%) / Teacher's Perspective											
	1.1 Intro & Pretest	1.2 Overview	2. Identifying a Problem	3. Locating Information	4.1 Evaluating Context	4.2 Evaluating Source	4.3 Evaluating Content	5. Synthesizing Information	6. Communictaing Findings	M (Case)	SD (Case)	Adherence Level by Teacher
Kasey	90.00	80.00	100.00	70.00	100.00	60.00	70.00	80.00	60.00	78.89	14.49	Moderate
Malena	80.00	90.00	80.00	60.00	60.00	60.00	60.00	60.00	60.00	67.78	11.33	Moderate
Amy (51)	100.00	90.00	100.00	90.00	100.00	80.00	80.00	80.00	80.00	88.89	8.75	High
Amy (55)	80.00	80.00	60.00	70.00	80.00	80.00	70.00		80.00	75.00	7.07	Moderate
Chris	90.00	60.00	100.00	100.00	90.00	90.00	100.00	90.00	80.00	88.89	11.97	High
M (Lesson)	88.00	80.00	88.00	78.00	86.00	74.00	76.00	77.50	72.00	79.89		Moderate

Note. Adherence levels: High 80% – 100%; Moderate 51% – 79%; Low 0% – 50% (An et al., 2020; Toomey et al., 2017.)

The following paragraphs contain the findings about adherence to content for individual cases, per the teacher's perspective. Only the most relevant information recorded in the Teacher's Journal regarding content adaptation is presented, per the evaluator's judgement. Minor changes recorded by the teachers are deemed 'normal' adjustments to the content.

Case 1. Kasey recorded a moderate adherence to content (78.9%) throughout the entire intervention. Kasey made one significant adaptation to the CORE intervention replacing the independent project for the 'Persuasive Paper' the class had been working in. This adaption impacted several lessons that included activities for the independent project. The lesson that Kasey found easier to teach was *Lesson 2 – Identifying a Problem.* Moreover, she adapted this lesson by bringing the Town hall Preparation activity to the beginning of the lesson, allowing her to expand the discussion on how students found the sources. Kasey scored *Lesson 6 – Communicating Findings*, with one of the lowest adherence to content scores (60%), justified by the fact that the students were familiar with the content. She suggested combining this lesson with *Lesson 5 – Synthesizing Information*. One other lowest score was assigned to *Lesson 4.2 – Evaluating Source* (60%), explained by her lack of preparation.

Case 2. Malena's perceived adherence to content was moderate (67.8%) throughout all the lessons. Like Kasey's case, she adapted the CORE intervention for her students to accommodate the 'Persuasive Paper.' Malena was the teacher who recorded the lowest score for adherence to content. During six of the nine lessons, Malena did not completely follow the lesson plan or did not use the provided material, recording a rating of 60%. According to her, she added some time for certain activities. She skipped others because she did not have time to complete them or change the activity to accommodate the student's contextual classroom setting. She mentioned that the examples should be more related to their regular class and incorporate the research they were doing for the speech. Regarding the Covid intervention topic, she believed it was too much. Like her colleague, Kasey, she recommended combining *Lesson 5 – Synthesizing Information* and *Lesson 6 – Communicating Findings*.

Case 3. Amy (51) recorded one of the highest adherence to content scores (88.9%) among the five cases. It is important to highlight that this was an ESL (English as a Second Language) class, so the intervention content has been adapted to include notes from Quebec's Progression of Learning at the Secondary Level English as a Second Language Core Programs. Amy (51) spent more time practicing intonation with the students and comparing it with 'French cognates.' Amy (51) adapted *Lesson 3 – Locating Information* in two different ways. First, she spent less time on the theory (less terminology), and second, she added an explanation of the terminology to the slides. That accommodation allowed for more time to prepare for the Town hall activity. Amy (51) was one of the teachers who decided to conduct the Independent Project as planned by the intervention creators. However, she crafted two new Word templates for students to complete the Independent Project's final delivery. Amy expressed that *Lesson 6 – Communicating Findings* was a topic her students had already studied during the academic year.

Case 4. Amy (55) reported a moderate adherence to content level (75%). This class is an ESL program with an adapted CORE intervention to comply with Quebec's Education

requirements. Amy (55) scored *Lesson 2 – Identifying a Problem*, with the lowest adherence score (60%), mainly explained by a timing issue. She started the lesson late and took time to explain and illustrate the different concepts contained in the lesson (e.g. perspective, beliefs, bias confirmation). Amy (55) did not have time to finish *Lesson 3 – Locating Information*. Since she had to postpone the Town hall activity for the following class, she assigned a score of 70% for adherence to content. Consequently, the adherence to content score was impacted for *Lesson 4.1 –Evaluating Context* recorded as 80%. This case highlights that Amy (55) conducted the GMO activity that was an extension of *Lesson 4.2 – Evaluating Source*, making her the only teacher to use the lesson extensions throughout the CORE program. Like Case 3, the class has already worked on the thesis concept, giving extra time to work on the Posttest and the Independent Project activities.

Case 5. Chris' class obtained a high level of adherence to content (88.9%). He assigned the lowest adherence score to *Lesson 1.2 – Overview,* mainly explained by the lack of Internet at the school at the time of the class; thus, he did improvise. Chris recommended moving the Town hall as the first activity of *Lesson 3 – Locating the Information*, to allow every student to express themselves without feeling hurried.

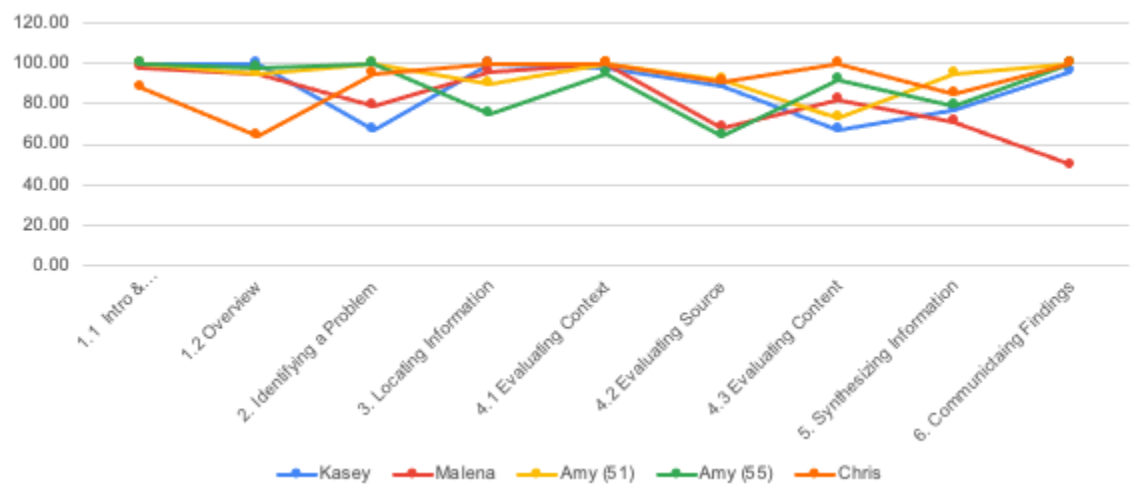
Adherence to Content – Researcher's Perspective

All teachers had high adherence to content level (89%) from the researcher's perspective (Figure 11, Table 7). The research observers assigned the highest level of adherence to Amy's (51) intervention (93.9%). This score might be explained by the fact that this was the second time the teacher delivered the intervention, unlike the other teachers who delivered only one instance of the program. On the contrary, Malena received the lowest adherence to content score (82.1%), most probably explained by the significant adaptation she made by combining *Lesson 5 – Synthesizing Information* and *Lesson 6 – Communicating Findings* into one lesson and reflected in that lessons scores of 71% and 50% respectively. The researchers reported that all the lessons had, on average, a high level of adherence above

80%, with *Lesson 4.1 – Evaluating Context* (self-study module) obtaining the best score (98.6%) and *Lesson 4.2 – Evaluating Source* the lowest score (80.8%). The high score for adherence to content for the self-study module might be self-explanatory, given that the teachers do not intervene in delivering the content but only provide instruction and guidance and fix accessibility issues. The lower score during *Lesson 4.2 – Evaluating the Source*, might be explained by student fatigue with the intervention and time limitation as recorded by the observer in the Observation Protocol.

Figure 11

Teacher Adherence to Content Scores – Researcher's Perspective



Illustrative Quotes:

"A lot of students are away on a school activity (ice fishing), she was not expecting this so she had to make a last-minute adjustment to her lesson plan [*Lesson 3 – Locating Information*]… students are unsure of what group they are supposed to be in… she says she cannot do it next week as she has to move on and she does not have time to do it next week" (Tristen).

"Most of the topics students proposed are unrelated to social justice. The teacher also had chosen to change the activity on slide 18 (original question was on should your school close during a pandemic?), she changed it to social justice related to connect to their other assignments she has them doing and the school's theme for the year" (Tristen).

"The teacher modified the lesson [*Lesson 5 – Synthesizing Information*] because this material was covered in previous lessons (prior to CORE project)" (Tristen).

"A lot of time is spent to solve access issues [*Lesson 4.1 – Evaluating the Context*]. Students are not able to navigate self-study on their own because they need access to videos and documents. We need to provide detailed instructions to teachers on how to make documents and links accessible to students" (Talya).

"Teacher went into town hall meeting before showing all of the websites. I think there were time concerns with this lesson's activities" (Ellen)

"There were several issues at the start of class, including that the student's didn't do their homework, that the projector wasn't working and the slides couldn't be projected and that they antibodies test had to be done during the class period. She also missed the previous class due to a snow day and the lesson 5 and 6 were blended. She was focusing on the speech the students need to do after spring break" (Ellen).

"Teacher added slides with the definitions for two reasons: 1-She said it was easier for her to teach the unit with the definition on screen (versus just being in the teacher notes) and 2-Teacher believed it was easier for the students to retain the information if they could read the definitions (versus just listening to her explain the terms without the definition up on screen)" (Kelly).

"The class will have a show tonight [*Lesson 4.2 – Evaluating Source*]. Amy conducted the lesson in a way not to get the students very tired. This lesson was finished on April 22nd, 2022" (Aria).

"The school did not have WIFI today, so the teacher had to improvise [*Lesson 1.2 – Overview*]. His solution was to write the questions and activity prompts on the board, and have the students look at the articles and their workbook on their phones. Some students did not have cellular data, so they could not participate in this activity" (Tristen).

Table 7

Teacher Adherence to Content Scores – Researcher's Perspective

Teacher	Total Adherence to Content by Lesson (%) / Researcher's Perspective											
	1.1 Intro & Pretest	1.2 Overview	2. Identifying a Problem	3. Locating Information	4.1 Evaluating Context	4.2 Evaluating Source	4.3 Evaluating Content	5. Synthesizing Information	6. Communictaing Findings	M (Case)	SD (Case)	Adherence Level by Teacher
Kasey	100.00	100.00	67.00	100.00	98.00	89.00	67.00	77.00	96.00	88.22	13.33	High
Malena	98.00	95.00	79.00	96.00	100.00	68.00	82.00	71.00	50.00	82.11	15.97	High
Amy (51)	100.00	95.00	100.00	90.00	100.00	92.00	73.00	95.00	100.00	93.89	8.21	High
Amy (55)	100.00	98.00	100.00	75.00	95.00	64.00	92.00	79.00	100.00	89.22	12.51	High
Chris	88.00	64.00	95.00	100.00	100.00	91.00	100.00	85.00	100.00	91.44	11.12	High
M (Lesson)	97.20	90.40	88.20	92.20	98.60	80.80	82.80	81.40	89.20	88.98	6.11	High

Note. Adherence levels: High 80% – 100%; Moderate 51% – 79%; Low 0% – 50% (An et al., 2020; Toomey et al., 2017.)

Teachers cut out several activities planned to be delivered in the CORE intervention. Table 8 displays the most skipped activities, including the Exit Ticket in *Lesson 1.2 – Overview*; the Bias Confirmation Reflection in *Lesson 2 – Identifying a Problem*; the Independent Project activity in *Lesson 4.2 – Evaluating Source*; the Summary in *Lesson 4.3 – Evaluating Content*

and the activity to Compare Articles in *Lesson 5 – Synthesizing Information*.

Table 8

Most Skipped Activities by all Teachers

Lesson	Part
1.2 Overview	Exit Ticket
2. Identifying a Problem	Bias Confirmation Reflection
4.2 Evaluating Source	Independent Project: Narrowing Down Resources
4.3 Evaluating Content	Summary
5. Synthesizing Information	Activity 2 Compare

Note. The criteria used to select the top five activities included activities skipped in at least three interventions.

The following paragraphs contain the findings about adherence to content for individual cases. It is important to note that only the most relevant information recorded in the Observation Protocol regarding teachers' content adaptation is presented, per the evaluator's judgement. Some slight changes reported by the observers are deemed ‘normal’ adjustments to the content.

Case 1. The research observers reported a high adherence to content score (88.2%) for Kasey’s class. The lowest adherence scores were assigned to *Lesson 2 – Identifying a Problem* (67%) and *Lesson 4.3 – Evaluating Content* (67%). One of the most relevant adaptations made by Kasey to the CORE program was the accommodation of the ‘Persuasive Paper’ as a replacement for the Independent Project. This adaptation led to skipping many of the related activities (e.g., Part 3, Selecting the Sources for the Independent Project during *Lesson 4.3 Evaluating the Content*). Kasey also skipped the Confirmation Bias Reflection in *Lesson 2. Identifying a Problem* and the Exit Ticket for *Lesson 1.2 – Overview*. An important factor to highlight during Kasey's intervention is the technical issue the teacher had (i.e., computer crashed) during the Pretest, reducing the time at task for this activity, and most probably affecting the intervention outcomes (i.e., the extent the students’ CORE skills improved

because of the intervention, addressed in a separate study conducted by the principal investigator.) Kasey faced difficulty delivering *Lesson 3 – Locating Information* because many students were absent (almost half the class) and did not know to which group they were assigned for the Town hall activity. The teacher adjusted the class without impacting the adherence to content scored as per noted by the observer. Throughout *Lesson 4.2 – Evaluating Source*, Kasey commented that students were tired of the topic and suggested fewer lessons in future interventions. During Lesson *4.3 – Evaluating Content*, students were instructed to complete the demographic questionnaire, thus reducing the time allocated to complete the lesson. In addition, Talya reported inconsistencies between the teachers' materials (presentation slides) and the student's workbook regarding the activity Preparation for Lesson 5 – *Synthesizing Information*, thus leading to the teacher's improvisation and postponing the activity for the subsequent lesson 5. Kasey modified *Lesson 5 – Synthesizing Information* because the content was previously taught. For example, Kasey "worked on concept mapping with the whole class instead of in pairs" (Talya). Instead of the Individual Concept Map activity, "the teacher asked students to complete a concept map for their individual essays" (Talya). During *Lesson 6 – Communicating Findings*, the teacher adjusted all activities of the Independent Project to the Persuasive Paper. Tristen said,

> Most of the topics students proposed are unrelated to social justice. The teacher also chose to change the activity on slide 18 (the original question was on should your school close during a pandemic?). She changed it to social justice related to connecting to their other assignments she has them doing and the school's theme for the year.

Case 2. The research assistants observed a high adherence to Malena's class content score (82.1%). The lowest adherence score was assigned to *Lesson 5 – Communicating Findings* (50%). This score could be mainly explained by Malena's major adaptation to the intervention, which blended *Lesson 5 – Synthesizing Information* and *Lesson 6 –*

Communicating Findings, into one lesson. In the end, the results are reflected in lower scores of content adherence and lack of time to deliver the entire material for these two lessons. In addition, Malena adapted the CORE program to fit the school's 'Persuasive Paper' project, thus skipping many of the related activities (e.g., Part 3 Selecting the Sources for the Independent Project during *Lesson 4.3 Evaluating the Content*). One important highlight is the technical issue students had while taking the Pretest, reducing the response time. *Lesson 4.1 – Evaluating the Context* presented many technical issues, such as navigation, access to videos and documents. Malena spanned the Town hall activity over two classes. There was insufficient time to cover all terminology in *Lesson 3 – Locating Information* (i.e., search engines, query syntax, filter bubbles, echo chambers). Malena modified *Lesson 4.2 – Evaluating the Source* to account for time limitations. For example, she walked the student through the video about how credible news organizations are instead of asking the students to evaluate their credibility by using the Jamboard-designed activity. Likewise, the teacher skipped some activities during *Lesson 4.3 – Evaluating Content*, such as skipping Fact-checking with Snopes activity and assigning as homework the reading of the articles in preparation for *Lesson 5 – Synthesizing Information*. Ellen portrayed this last lesson 5 as being disrupted by several circumstances (e.g., the projector was broken, antibodies tests delivery, missing previous lesson) that affected the proposed flow of the lesson, thus skipping activities (e.g., activate prior knowledge on summary and synthesis, concept map example) or giving activities for homework (e.g., Part 2 organize articles that students had to read before class).

Case 3. The research assistants observed a high adherence to content score (93.9%) for Amy (51)'s class. Amy (51) made a few minor adaptations to the CORE program to accommodate timing and specific class circumstances. The teacher postponed many activities when she could not finish during the planned lesson. However, the teacher allocated extra time after the last lesson to let all students present the Independent Project (these extra classes were not observed). The lowest adherence score was assigned to *Lesson 4.3 – Evaluating*

Content (73%), mainly explained by an engagement factor (i.e., the students had an important physics exam the next period), motivating the teacher to adapt the lesson to this specific context condition. It is important to note that Amy (51) is in an ESL class, and the teacher focused throughout all lessons on the English vocabulary and where to put the stress. Amy (51) adapted the intervention material by adding the definitions to the slides. During *Lesson 3 – Locating Information*, Amy (51) did not have time to complete the Town hall activity. Hence, during l*esson 4.1– Evaluating the Context*, the teacher took the first 20 minutes of the class to finish the town hall presentations.

Case 4. The research observers scored high adherence to content (89.2%) to Amy (55)'s intervention. The lowest score was granted to *Lesson 4.2 – Evaluating Source* (64%), mainly because the students had a talent show to perform that night. This contextual event led the teacher to skip some activities (e.g., Types of Organizations and impact on its content credibility), change the activity approach (e.g., individual activity changed to group activities, interactive activity changed to direct instruction), or postpone for the next class. Amy was the only teacher to take advantage of the 'extended activities' and played the NON-GMO project video in both groups 51 and 55, modelling the credibility evaluation process by an expert using the think-aloud method. *Lesson 6 – Communicating Findings* was used by the teacher to allow the student to complete the demographic questionnaire and to hand over some consent/assent forms. This activity did not disrupt the lesson content since students had previous knowledge of the topic (i.e., writing a thesis statement), thus allowing the teacher to go quickly through the content and even skip some activities.

Case 5. The research observers scored high adherence to content (91.4%) to Chris' intervention. The lowest score was granted during *Lesson 1.2 – Overview* (64%), mainly explained by the lack of Internet at the school and the teacher's improvisation. One of the major adaptations Chris introduced to the CORE program was the Independent Project, which was related to his history class mandate: "Ten days that shook the world." Throughout the entire

intervention, Chris adapted or extended the examples to relate them to the history class. In addition, when the activities were not done during the class period, Chris assigned them as homework (e.g., Exit Slip). It is important to note that Chris extended the CORE intervention to following class periods to allow students to prepare their CORE assignment (i.e., Independent Project). However, the researchers did not observe these extra classes. During *Lesson 2 – Identifying a* Problem, Chris adapted the presentation slides to add the Confirmation of Bias concept in more detail. Another adaption Chris had to make was to remove some of the roles for the Town hall activity given the smaller size of his class, which was an intended adaptation within the lesson design. During *Lesson 4.1 – Evaluating Context*, students had many issues with access, video, and audio. These technical difficulties triggered the teacher to adapt the approach from a self-study module to "having to do the activity together as a class," as stated by Tristen. Students in Chris' class had previous knowledge of the concept maps, covered by *Lesson 5 – Synthesizing Information*, driving the teacher to make some adaptations, such as blending the concept map general activity with example A.

Research Question 2: Quality of Delivery

Quality of delivery is a potential moderator of the relationship between the intervention and the program implementation fidelity. It is defined as how the implementer delivers the program, thus using the prescribed techniques, processes, and methods. However, it was operationalized for our study by measuring teachers' preparedness, confidence, and enthusiasm to deliver the content. If the content of an intervention is badly delivered, then the implementation fidelity level might be negatively affected.

Overall, teachers had a high-quality delivery score from a consolidated perspective, 93.3% from teacher's perspective and 96.7% from researcher's view, as shown in Figure 12, Table 9. Therefore, we can argue that quality of delivery positively impacted the implementation fidelity of the CORE program. Amy (51) scored the highest quality of delivery average (99.3%), while Malena achieved the lowest score (79.5%). This latter score was mainly driven by the

teacher self-assessment (67.4%), which in turn is explained by the teacher being unfamiliar with the content and the intervention terminology, according to the notes logged in the Teacher's Journal. An important hint to remember for the next program iteration is teachers' training and preparation.

Comparatively, the researcher's perspective regarding delivery quality was higher (96.7%) than the teachers' (90.3%). In the cases of Kasey, Malena and Amy (55), the scores for quality of delivery given by the researchers were higher than the teacher's perspective (see Figure 12). Conversely, Amy (51) and Chris perceived their quality of delivery as higher than the reported by the researchers; however, these differences were minimal (less than 2%). Malena's case reported the highest difference between the researcher and teacher perspectives (24.3%) in the scores for quality of delivery. This difference could probably be explained by Malena's self-reported lack of confidence in delivering some complex concepts of the CORE lessons, such as the terminology in *Lesson 3 – Locating Information*, which the research observers might not noticed. On the contrary, Amy (51) reported the slightest difference (-1.3%) between these two perspectives.

Figure 12

Quality of Delivery by Perspective

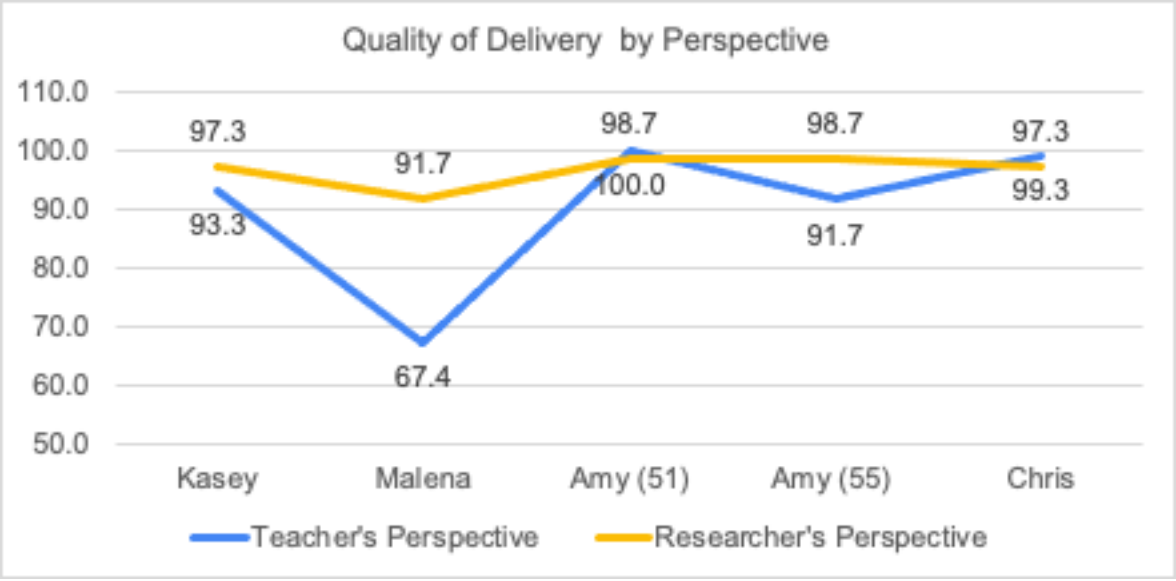


Table 9

Consolidated Quality of Delivery Scores

Teacher	Consolidated Quality of Delivery Score (%)		
	Teacher's Perspective	Researcher's Perspective	Difference between Perspectives
Kasey	93.33	97.33	4.00
Malena	67.41	91.67	24.26
Amy (51)	100.00	98.67	-1.33
Amy (55)	91.67	98.67	7.00
Chris	99.26	97.33	-1.93
M (Perspective)	90.33	96.73	6.40
Total Quality Level by Perspective	High	High	

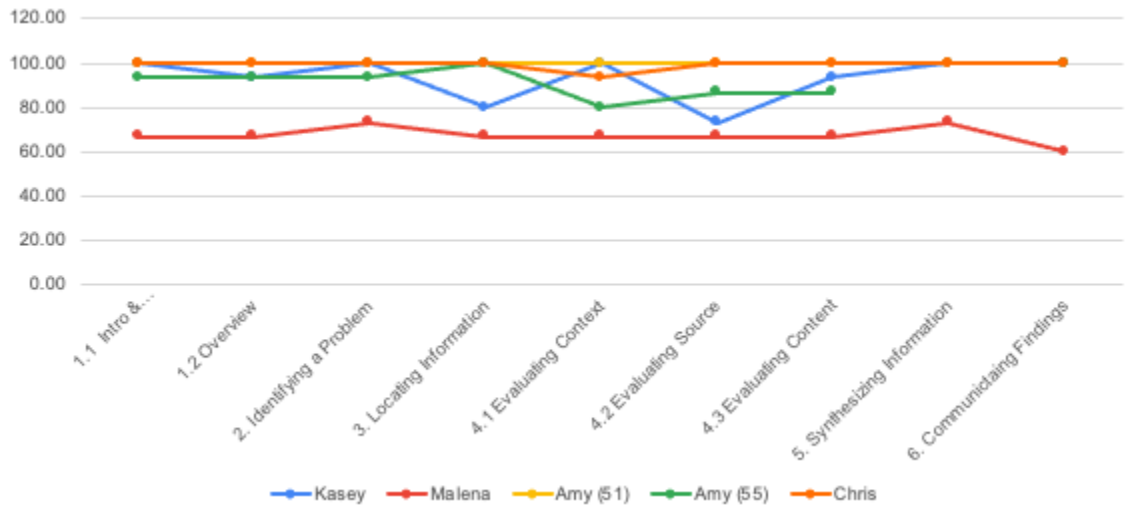
Note. Adherence levels: High 80% – 100%; Moderate 51% – 79%; Low 0% – 50% (An et al., 2020; Toomey et al., 2017.)

Quality of Delivery – Teacher's Perspective

Teachers had a high-quality delivery level (90.3%) from their perspective (Figure 13, Table 10). Amy (51) reported the highest levels of quality (100%) throughout all the lessons, while Malena was the only teacher reporting a moderate level of quality of delivery (67.4%). The teachers reported *Lessons 2 – Identifying a Problem* as the lesson they perceived they delivered with the most confidence and preparation, assigning a score of 93.3%. Conversely, they reported *Lesson 4.2 – Evaluating Source*, with the lowest score in confidence and preparation (85.3%.) Teacher noted certain aspects that might explain the lower quality of delivery scores for lesson 4.2, such as struggle toggling between the presentation slides and the student's workbook, the difficulty with using Jamboard technology, and the complexity of the concepts. As a corrective measure to some of the mentioned issues, Amy (51) expressed that she modified the content of the presentation slides after going through the intervention with the previous classroom (55).

Figure 13

Teacher Quality of Delivery Scores – Teacher's Perspective



Illustrative Quotes:

"Despite having read the work over in advance I find it hard to navigate between the workbook and the slides. The info on the slides would also be good in the workbook. I felt somewhat uneasy with presenting the information (this was a first try) [*Lesson 4.2 – Evaluating the Source*]. Each time I do an activity it is easier the second try as I become more familiar" (Kasey).

"The technical terms [*Lesson 3 – Locating Information*] made it difficult… Some of the information [*Lesson 4.1 – Evaluating Context*] I am not familiar with so it makes it difficult to teach…Some of the terms were difficult [*Lesson 4.2 – Evaluating Source*] …I think more info for the teachers on the key terms [*Lesson 4.3 – Evaluating Content*] would be helpful" (Malena).

"I felt confident after having taught the same lesson yesterday, I no longer needed to read my notes and the teacher's notes on the slides before the class. I also anticipated potential technical problems, after yesterday's class" (Amy [51]).

"I did not feel prepared before the class [*Lesson 3 – Locating Information*], because there were many new terms, for me and for the Ss [students]. I read all T's [teacher's] notes and I decided to modify the slides to add the definitions, examples and details so Ss could have all necessary information at hand. I also added the difference between Internet and World Wide Web and other examples of how punctuation can improve query syntax. I added or changed some images from the original slides (some were not well superposed.) … I felt confident once I had added all the info and visual support" (Amy [55]).

"I felt prepared for the lesson [*4.1– Evaluating the Context]* , but the last-minute change due to access did make it shaky at points" (Chris).

Table 10

Teacher Quality of Delivery Scores – Teacher's Perspective

Teacher	Total Quality of Delivery Score by Lesson (%) / Teacher's Perspective											
	1.1 Intro & Pretest	1.2 Overview	2. Identifying a Problem	3. Locating Information	4.1 Evaluating Context	4.2 Evaluating Source	4.3 Evaluating Content	5. Synthesizing Information	6. Communictaing Findings	M (Case)	SD (Case)	Quality Level by Teacher
Kasey	100.00	93.33	100.00	80.00	100.00	73.33	93.33	100.00	100.00	93.33	9.43	High
Malena	66.67	66.67	73.33	66.67	66.67	66.67	66.67	73.33	60.00	67.41	3.78	Moderate
Amy (51)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	0.00	High
Amy (55)	93.33	93.33	93.33	100.00	80.00	86.67	86.67		100.00	91.67	6.45	High
Chris	100.00	100.00	100.00	100.00	93.33	100.00	100.00	100.00	100.00	99.26	2.10	High
M (Lesson)	92.00	90.67	93.33	89.33	88.00	85.33	89.33	93.33	92.00	90.33		High

Note. Quality of delivery levels: High 80% – 100%; Moderate 51% – 79%; Low 0% – 50% (An et al., 2020; Toomey et al., 2017.)

The following paragraphs contain the findings about the quality of delivery for individual cases. It is important to note that only the most relevant information recorded in the Teacher's Journal regarding confidence and preparedness is presented, per the evaluator's judgement.

Case 1. Kasey recorded a high-quality delivery score (93.3%) throughout the entire intervention. The lowest score on the quality of delivery was reported for *Lesson 4.2 – Evaluating the Source,* mainly explained by the difficulty of navigating between the slides and the workbook. Kasey expressed her uneasiness in conducting the lesson for the first time. She mentioned that it is always easier the second time she does an activity. Kasey mentioned two important remarks related to confidence and preparation. First, she found the Jamboard technology "hard to work with," impacting her confidence in teaching the lessons. Secondly, during *Lesson 3 – Locating information*, she did not feel prepared due to the complexity of the concepts. She said the concepts were difficult to explain and required more explanations.

Case 2. Malena self-reported a moderate level of quality delivery (67.4%). In the Teacher's Journal, Malena expressed the need for more "teacher instruction on how to carry out

lessons.” In general, she did not feel familiar with the technical terminology and concepts, making it challenging for her to teach.

Case 3. Amy (51) self-reported a very high level of quality of delivery (100%) conducting the full intervention, mainly driven by the fact that this was her second time delivering the lessons, and she felt more confident than the first time. In addition, Amy (51) was able to anticipate some technical issues. One important observation is that after teaching the lesson for the first time in the other group (55), Amy (51) modified some of the slides. For example, during *Lesson 4.2 – Evaluating the Source*, she said, “I put all the author slides together and I added slides from a previous lesson to reminder of how to read laterally and fact-checking websites” (Aria).

Case 4. Amy (55) self-reported a high level of quality of delivery (91.7%). Amy made explicit in the Teacher’s Journal that she had prepared the lesson in advance and read all of the teacher’s instructions. Moreover, she adapted the lessons' material to accommodate her teaching style, class context, and self-confidence. For example, for *Lesson 3 – Locating Information*, she read the material in advance and modified the slides with extra information she deemed important in the presentation.

Case 5. Chris reported a very high level of quality of delivery (99.3%) through the intervention. He mentioned to have read over the lesson in advance. The lesson where he had a small challenge was *4.1– Evaluating the Context* because although he felt prepared, access issues made it somewhat weak.

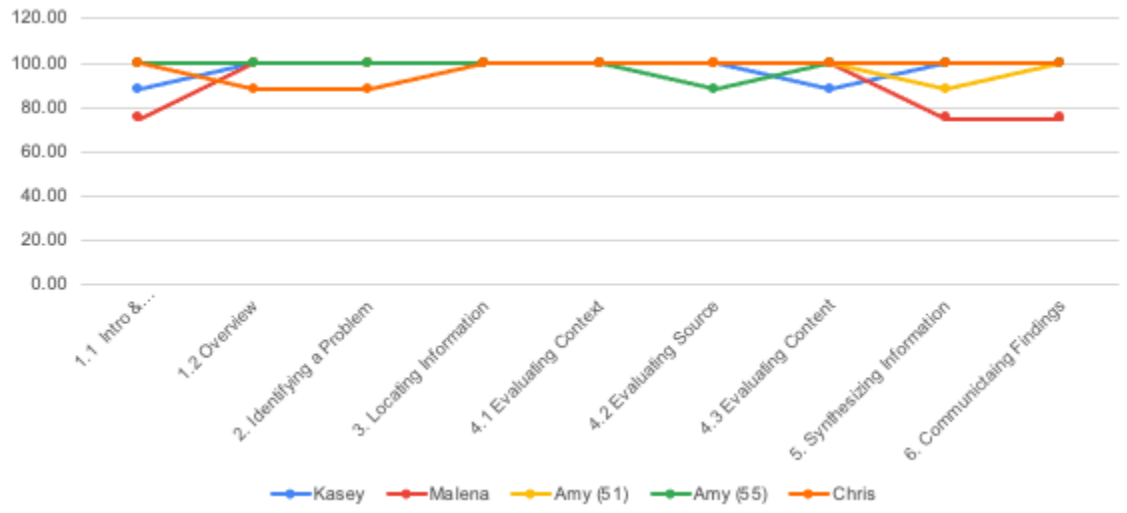
Quality of Delivery – Researcher’s Perspective

All teachers had a very high quality of delivery level (96.7%) from the researcher’s perspective (Figure 14, Table 11) throughout all the lessons. The researchers reported that all the lessons had, on average, a high level of quality above 90%, with *Lesson 3 – Locating Information* and Lesson *4.1 – Evaluating Context* (self-study module) obtaining a ‘perfect’ score of 100%. The high score for quality of delivery for the self-study module might be self-

explanatory, as it was for adherence to content, given that the teachers do not intervene in delivering the content but only provide instruction and guidance and fix accessibility issues. On the contrary, the perfect score for *Lesson 3 – Locating Information* is challenging to explain, given that all evidence recorded in the Teacher's Journal revealed some teachers feeling apprehensive and nervous due to the difficulty and complexity of the concepts for them to deliver. Nevertheless, observers did not notice any teachers' self-doubts. The evidence in the notes recorded in the Observation Protocol suggests a high level of preparation for this lesson, offering a plausible explanation for these extremely high scores.

Figure 14

Teacher Quality of Delivery Scores – Researcher's Perspective



Illustrative Quotes:

"The teacher was very experienced and was well prepared to teach the lesson. She had said that she felt a bit nervous because it was her first time teaching this particular lesson [*1.1 – Intro and Pretest*]. She asked alot of CCQs which kept the students engaged. The students had good rapport with this teacher and it was obvious that there was a positive relationship between teacher and students" (Tristen).

"The teacher commented that she found it easier to deliver the previous lessons, but she was nervous to deliver this [*Lesson 3 – Locating Information*] as it is very heavy with lots of terms" (Tristen).

"[Malena] did a great job of engaging the ss [students] and giving more explanations when they were needed" (Ellen).

"The teacher uses tactics to engaged students. One of the students I was observing became engaed trhough questions by the teacher" (Aria).

"The teacher did a really good summary of the program as of today [*Lesson 4.2 – Evaluating Source*]. She included the importance of this lesson during the CEGEP and university" (Aria).

"The teacher was very prepared for the class [*Lesson 3 – Locating Information*] and went over how to use Snopes.com on slide 17 in great detail, giving a demonstration. Overall, the teacher has a good, formal relationship with the students and communicated high expectations for their behaviour and achievement" (Ellen)

Table 11

Teacher Quality of Delivery Scores – Researcher's Perspective

Teacher	Total Quality of Delivery by Lesson (%) / Researcher's Perspective											
	1.1 Intro & Pretest	1.2 Overview	2. Identifying a Problem	3. Locating Information	4.1 Evaluating Context	4.2 Evaluating Source	4.3 Evaluating Content	5. Synthesizing Information	6. Communictaing Findings	M (Case)	SD (Case)	Quality Level by Teacher
Kasey	88.00	100.00	100.00	100.00	100.00	100.00	88.00	100.00	100.00	97.33	4.99	High
Malena	75.00	100.00	100.00	100.00	100.00	100.00	100.00	75.00	75.00	91.67	11.79	High
Amy (51)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	88.00	100.00	98.67	3.77	High
Amy (55)	100.00	100.00	100.00	100.00	100.00	88.00	100.00	100.00	100.00	98.67	3.77	High
Chris	100.00	88.00	88.00	100.00	100.00	100.00	100.00	100.00	100.00	97.33	4.99	High
M (Lesson)	92.60	97.60	97.60	100.00	100.00	97.60	97.60	92.60	95.00	96.73	2.61	High

Note. Quality levels: High 80% – 100%; Moderate 51% – 79%; Low 0% – 50% (An et al., 2020; Toomey et al., 2017.)

The following paragraphs contain the findings about the quality of delivery for individual cases from the researchers perspective. Of importance to note, only the most relevant information recorded in the Observation Protocol regarding the teacher's confidence and preparedness is presented, per the evaluator's judgement.

Case 1. Research observers assigned a high-level score (97.3%) to quality delivery for Kasey, revealing the teacher's good preparation and confidence in teaching the intervention. According to Ellen, the teacher mentioned being nervous about teaching the class for the first time. *Lesson 3 – Locating Information* triggered apprehension in Kasey due to the complexity of the terminology and concepts. However, from the researcher's perspective, Kasey obtained a

high score of 100% for quality delivery for Lesson 3, most probably derived from preparation and confidence demonstrated in delivering the content. The observers jotted down several teacher behaviours that demonstrated her preparation and confidence. For example, the use of ‘concept checking questions (CCQs)’ for checking students’ comprehension, motivating quiet and silent students, ‘with-it-ness' to perceive the needs of her students with accuracy and care, surveying the class for understanding, directed questions at particular disruptive on disengaged students, making the topic ‘relatable’ with personal experiences. Tristen observed, “Teacher seemed highly interested in the subject and, by sharing examples from her own life, encouraged students to take the topic seriously.”

Case 2. The research observers recorded a high score (91.7%) for the quality of delivery for Malena’s case. According to the observers, the teacher demonstrated preparation and confidence behaviours, such as: setting Google Classroom for the intervention, writing guiding questions on the board, pointing students to the glossary to help students identify and acquire the vocabulary of the CORE intervention, explaining the main idea of the videos after playing them, asking students to share their opinions.

Case 3. The research observers granted Amy (51) a very high delivery of quality score (98.7%). According to the researchers, the teacher demonstrated preparation and confidence through the behaviours while conducting the intervention. For example, setting up in advance the Showbie application to assign, collect, and review student work (i.e., Student’s Workbook, independent project), making concepts relatable and usable in day-to-day activity, highlighting the importance of CORE to their near academic future in CEGEP, redesigning some slides according to her teaching style and her classroom needs, handling disruptive students through direct questions, or asking to stop any non-related CORE tasks, using examples from social media platforms used by young students (e.g., Instagram and Tik Tok.)

Case 4. According to the research observers, Amy (55) demonstrated a high level of quality of delivery (98.7%). The behaviours revealed a high preparation and confidence to

conduct the intervention. In addition to the practices noted in see previous case 3, Amy (55), used to write down the agenda on the board, presented suitable examples such as Google Scholar, repeated instructions when appropriate, motivated students to write in the student's workbook, emphasized on the importance of the CORE program.

Case 5. The research observers logged a high level of quality of delivery (97.3%) for Chris' intervention, demonstrating preparation and confidence behaviours during the CORE intervention. For example, Chris managed disruptive students (either by asking them to stop talking or separating groups when they were not focused on the CORE tasks), reminding the students to be quiet, walking around the classroom to check students' focus, adapting the CORE lesson examples to his subject class (i.e., History), walking the students through the activities and offering clarifications when required, calling out students by name to answer questions. According to Ellen, "The teacher was very prepared for the class and used an engaging, communicative style with the students." Chris explicitly told the students about his high expectations of behaviours and achievements.

Research Question 3: Participant Responsiveness (Engagement and Motivation)

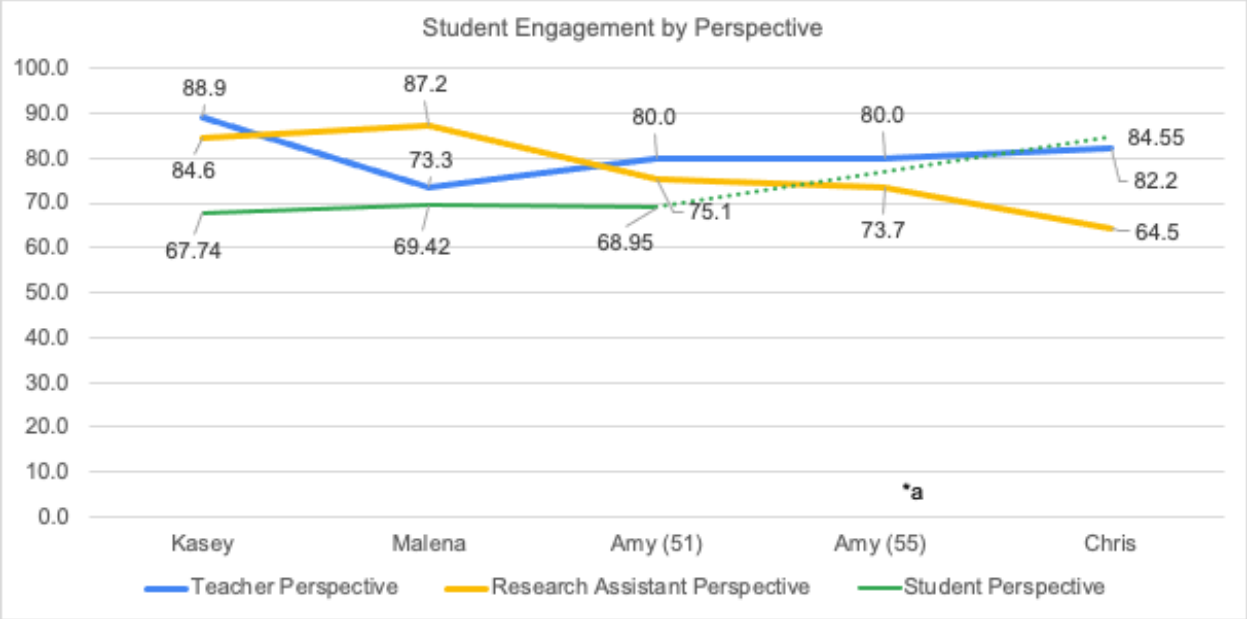
Participant responsiveness refers to the extent to which participants are engaged by and involved in the activities and content of the intervention. It includes participants' judgement about the outcomes and relevance of the program. Participant responsiveness is a moderating variable of implementation fidelity. If participants perceived the intervention as irrelevant or unimportant, it might negatively affect implementation fidelity. We operationalized participant responsiveness as how engaged and motivated the students were during the intervention.

Students' engagement and motivation varied from the different views. From the teacher's perspective, students were highly engaged, with a mean score of 80.9%, while researchers and students perceived they were moderately engaged and motivated (77%, 69.6%), as shown in Figure 15, Table 12. Only one case, Kasey, achieved a high-level score of 80.5%, mainly driven by the teacher's high perception of her student's engagement and motivation (88.9%.) The rest

of the cases had a moderate level of student engagement and motivation, above 74.5%.

Figure 15

Consolidated Student Engagement Scores



[a] Data not available

Table 12

Consolidated Student Engagement Scores

Teacher	Consolidated Student Engagement Score (%)					
	Teacher Perspective	Research Assistant Perspective	Student Perspective	Difference between Teacher and Student	Difference between Researcher and Student	Difference between Teacher and Researcher
Kasey	88.89	84.60	67.74	21.15	16.86	4.29
Malena	73.33	87.19	69.42	3.91	17.76	-13.85
Amy (51)	80.00	75.12	68.95	11.05	6.17	4.88
Amy (55)	80.00	73.67	-[a]	-[a]	-[a]	-[a]
Chris	82.22	64.47	84.55	-2.32	-20.07	17.75
M (Perspective)	80.89	77.01	72.66			
Total Engagement Level by Perspective	High	Moderate	Moderate			

Note. Engagement levels: High 80% – 100%; Moderate 51% – 79%; Low 0% – 50% (An et al., 2020; Toomey et al., 2017.)

[a] Data not available

Working in groups, interactivity, and allowing to share ideas and perspectives are highlighted as one of the most engaging and motivating factors through the entire intervention, across all cases, and from the three perspectives (teacher's, researcher's, and student's). On

the contrary, the length of the program, repetition, and the feeling of already knowing most of the information are the characteristics that were highlighted as the least engaging and motivating aspects.

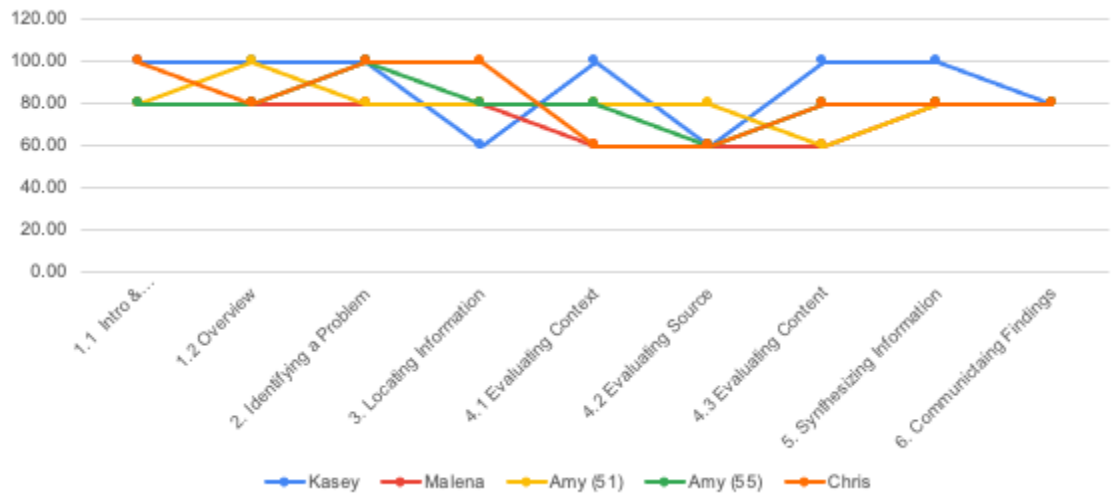
Three cases reported student self-assessment scores below the teacher's (Kasey 21%, Malena 4%, and Amy (51) 11.5%). Likewise, students' self-assessment was below the researcher's perspective (Kasey 16.9%, Malena 17.8%, and Amy (51) 6.6%). Only Chris' case reported higher student engagement and motivation scores than teachers (-2.3%) or researchers (-20.1%). The differences between the teacher's and the researcher's perspectives were inconsistent across the five cases (Kasey 4.3%, Malena -13.9%, Amy (51) 4.9%, and Chris 17.8%). Student engagement and motivation are complex constructs in education and psychology. They encompass various factors and are influenced by multiple internal and external elements. The reasons that might explain the abovementioned differences were not evident in the collected data. Nonetheless, the different data collection instruments might partially clarify the results. For example, the Student Survey was a self-reported instrument that captured data once at the end of the intervention. In contrast, the Observation Protocol captured scores from external researchers throughout all the lessons. At the end of the CORE program, students were tired and bored, which might have influenced a lower score for some teachers.

Participant Responsiveness – Teacher's Perspective

Students had a high engagement and motivation level (80.9%) from a consolidated teacher's perspective (Figure 16, Table 13). In four cases, teachers scored students' engagement and motivation as high (above 80%), and only Malena perceived these behaviours as moderate (73.3%). During most lessons, teachers reported high student engagement and motivation levels, with *Lesson 2 – Identifying a Problem* attaining the top score (92%). Conversely, *Lesson 4.2 – Evaluating Source* was perceived by all teachers as the less engaging class with a 64% engagement score. Details about findings by lesson will be presented further in the *Student Engagement Scores – Researcher's Perspective* section.

Figure 16

Student Engagement Scores – Teacher's Perspective



Illustrative Quotes:

"Concepts took long to explain [*Lesson 3 – Locating Information*.] This was a very heavy lesson for concepts. Perhaps hard for students to follow along. Lesson had to be rushed to complete because the terms were hard to explain" (Kasey).

"Students are seemingly getting 'fatigued' by the lessons, they find themselves losing interest (perhaps too many lessons on the topic). They remain engaged but I feel a sense that it is too much each class. I changed the concept map idea [Lesson – 5 Synthesizing Information] a bit as students are getting bored so I chose to focus on their own persuasives instead" (Kasey).

"These lessons are getting repetitive and could be combined to make it shorter. Losing interest talking about Covid. I am going to move away from Covid and start focusing on their own research" (Malena).

"Ss [students] were more engaged during the Town Hall preparation and very engaged during the debate. They presented excellent arguments supported by good sources…the interactive voting improved their engagement" (Amy [51]).

"The first part was long, I had to elicit examples and insist on the importance of understanding why we are in a Filtre Bubble when on social media. They were more engaged during the Town Hall activity" (Amy [55]).

"Students were not overly enthused with this topic and the self-study as it did not have a lot of interactions from their perspective…Students needed to be prompted to remain on task. They seemed disinterested and confused during some of the activities due to the heavy amount of lecturing at the beginning" (Chris).

Table 13

Student Engagement Scores – Teacher's Perspective

Teacher	Total Student Engagement Score by Lesson (%) / Teacher's Perspective											
	1.1 Intro & Pretest	1.2 Overview	2. Identifying a Problem	3. Locating Information	4.1 Evaluating Context	4.2 Evaluating Source	4.3 Evaluating Content	5. Synthesizing Information	6. Communictaing Findings	M (Case)	SD (Case)	Engagement Level by Teacher
Kasey	100.00	100.00	100.00	60.00	100.00	60.00	100.00	100.00	80.00	88.89	16.63	High
Malena	80.00	80.00	80.00	80.00	60.00	60.00	60.00	80.00	80.00	73.33	9.43	Moderate
Amy (51)	80.00	100.00	80.00	80.00	80.00	80.00	60.00	80.00	80.00	80.00	9.43	High
Amy (55)	80.00	80.00	100.00	80.00	80.00	60.00	80.00	80.00	80.00	80.00	9.43	High
Chris	100.00	80.00	100.00	100.00	60.00	60.00	80.00	80.00	80.00	82.22	14.74	High
M (Lesson)	88.00	88.00	92.00	80.00	76.00	64.00	76.00	84.00	80.00	80.89		High

Note. Engagement levels: High 80% – 100%; Moderate 51% – 79%; Low 0% – 50% (An et al., 2020; Toomey et al., 2017.)

The following paragraphs contain student engagement and motivation findings for individual cases. Only the most essential information recorded in the Observation Protocol about students' engagement and motivation is presented, per the evaluator's judgement.

Case 1. Kasey perceived that her students were generally engaged and motivated during the intervention (88.9%). Students were less engaged during *Lesson 3 – Locating Information*(60%) and *Lesson 4.2– Evaluating Source* (60%). During *Lesson 3 – Locating Information*, there were many absent students, making the Town Hall activity difficult to finish. According to Kasey, this lesson was complex and challenging for students. The concepts were heavy, and the time to explain them was longer than expected. Kasey observed students were getting tired of the intervention by *Lesson 4.2 – Evaluating Source*. She changed and adapted a few lessons to keep students interested.

Case 2. Malena observed that her students had moderate engagement and motivation throughout the intervention (73.3%). She noted that students were mainly engaged when working in groups or on their topic, such as their speech's introduction and thesis statement.

Similarly to Kasey's group, students were fatigued by *Lesson 4.2 – Evaluating Source*, perceiving the program as repetitive and uninterested in listening or talking about Covid.

Case 3. Amy (51) noted that engagement was high for most of the intervention (80%). This group "tends to be more studious but less talkative." According to Amy (51), keeping students engaged during the theory for *Lesson 3 – Locating Information was difficult*. Engagement increased during the Town Hall activity, and the adaptation Amy made by incorporating an 'interactive voting' was successful in keeping students interested. *Lesson 4.3 – Evaluating Content* was perceived by Amy (51) as the less engaging topic for her students (60%). She partially explained it through an upcoming physics exam. As perceived by other teachers, up to this point of the intervention, students started to feel the intervention was 'repetitive.' According to Amy (51), students were highly engaged with the activities related to the Independent Project, looking for sources, making the concept map, the synthesis and preparing for the presentation.

Case 4. Amy (55) noticed the student's high engagement and motivation during almost the entire intervention (80%). *Lesson – 4.2 Evaluating Context* was the class where students were more distracted and tired, mainly explained by a big art show students had to perform that night and the following night. Amy adapted the lesson to be more teacher-led, given the specific situation of the class. *Lesson – 3 Locating information* was long, as previously mentioned by other teachers. Similar to Case 3, students denoted more interest and engagement during the Town Hall activity.

Case 5. Chris perceived a high level of engagement and motivation throughout the entire intervention (82.2%). There were two lessons with the lower engagement score: *4.1 – Evaluating Context* (60%) and *4.2 – Evaluation Source* (60%). Chris mentioned the students' lack of enthusiasm for the topic and the self-study.

Participant Responsiveness – Researcher's Perspective

Students had a moderate (77%) engagement and motivation level from the researcher's perspective (Figure 17, Table 14). Malena and Kasey obtained the top scores (87.2% and 84.6%, respectively). The first three lessons of the program, Lesson *1.1– Intro*, Lesson *1.2 – Overview*, and Lesson *2 – Identifying a Problem,* reported the highest student engagement and motivation levels, with the introductory lesson having the top percentage (92.8%). The lowest score for student engagement recorded by the researchers was for *Lesson 3 – Locating Information* (67.1%). The highest score might be explained by the excitement of initiating the CORE program and the engaging activities proposed in the lesson content. The lowest score for *Lesson 3 – Locating Information* is likely explained by the difficulty and complexity of the concepts, as already known by the findings for the adherence to content and quality of delivery dimensions.

Figure 17

Student Engagement Scores – Researcher's Perspective

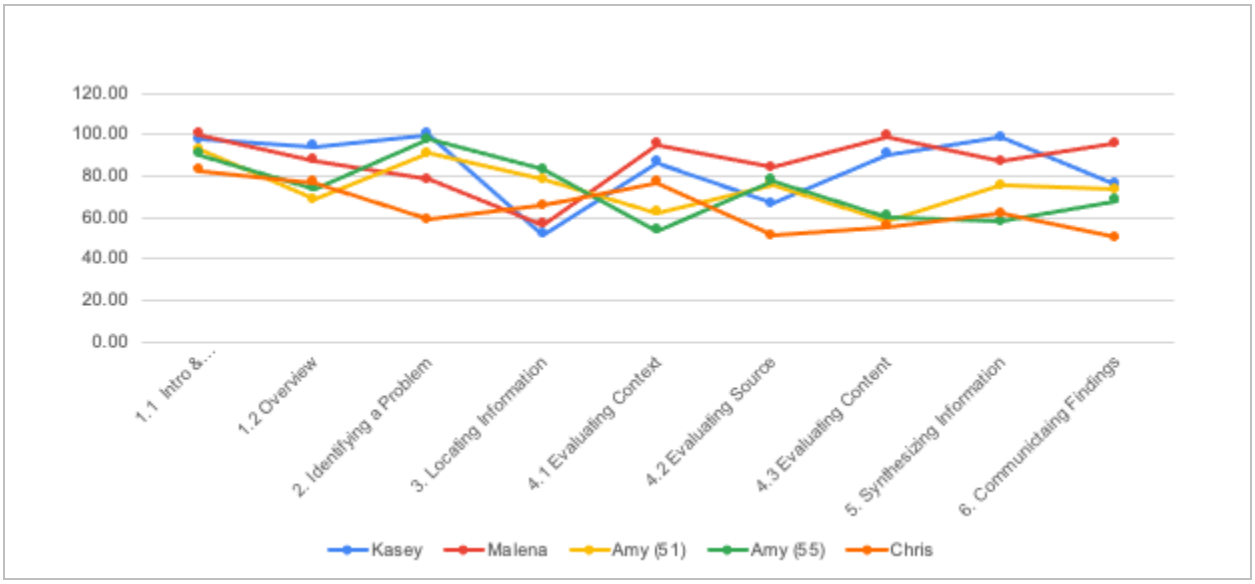


Table 14

Student Engagement Scores – Researcher's Perspective

Teacher	Total Student Engagement Score by Lesson (%) / Researcher's Perspective											
	1.1 Intro & Pretest	1.2 Overview	2. Identifying a Problem	3. Locating Information	4.1 Evaluating Context	4.2 Evaluating Source	4.3 Evaluating Content	5. Synthesizing Information	6. Communictaing Findings	M (Case)	SD (Case)	Engagement Level by Teacher
Kasey	97.62	94.45	100.00	51.63	86.67	66.55	90.32	98.49	75.70	84.60	15.72	High
Malena	100.00	87.79	78.57	56.46	95.37	84.29	99.42	86.95	95.84	87.19	12.84	High
Amy (51)	93.10	68.43	90.78	78.19	62.50	76.06	58.27	75.39	73.33	75.12	10.90	Moderate
Amy (55)	90.21	73.75	97.50	83.33	53.75	77.67	60.49	58.00	68.28	73.67	14.13	Moderate
Chris	82.82	76.69	59.19	65.96	77.05	51.19	55.68	61.66	50.00	64.47	11.28	Moderate
M (Lesson)	92.75	80.22	85.21	67.11	75.07	71.15	72.83	76.10	72.63	77.01	7.44	Moderate

Note. Engagement levels: High 80% – 100%; Moderate 51% – 79%; Low 0% – 50% (An et al., 2020; Toomey et al., 2017.)

Activities with Highest Student Engagement Scores. The top three activities that most effectively engaged students throughout all five interventions were: *Introduction to the CORE Program*, with the highest score of 94.6%, followed by *Beliefs Self-Assessment* (91.1%), and *Incredible Images* (88.1%), as displayed in Table 15. On the contrary, the least three engaging activities: *Writing a Thesis Statement* (60.6%), followed by *Intro to Locating Information* (60.7%), *and Lateral Reading* (64.5%.), as seen in Table 16. There are a variety of reasons that might explain student engagement and motivation during the top ten scored activities, as well as the least engaging and motivating activities. The evidence suggests that students were more engaged and motivated during the initial lesson, representing seven of the ten top-scored activities. One more common factor is the type of activity, which mostly involves group discussion and participation. Likewise, the least engagement activities were towards the end of the intervention when students felt 'intervention fatigue,' some of the content they already knew, or the lesson was heavy in complex terminology ad direct instruction.

Note. The full details on student engagement scores by activity are shown in Appendix B.

Table 15

Top 10 Activities with Highest Student Engagement Scores – Researcher's Perspective

Lesson	Part	Engagement Score by Activity (Researcher's Perspective)					
		Kasey	Malena	Amy (51)	Amy (55)	Chris	M
1.1 Intro & Pretest	Introduction to the CORE program	92.86	100.00	100.00	92.86	87.50	94.64
2. Identifying a Problem	Beliefs Self Assessment	100.00	100.00	96.67	100.00	58.82	91.10
1.2 Overview	Warm-Up: Incredible Images	100.00	76.00	76.47	100.00		88.12
1.1 Intro & Pretest	Warm-Up: Pizza Indicators	100.00	100.00	79.31	77.78	80.95	87.61
4.3 Evaluating Content	Lesson 5 Preparation	100.00	100.00	100.00		42.86	85.72
1.2 Overview	Looking for Indicator Activity	100.00	91.67	86.21	50.00	100.00	85.58
2. Identifying a Problem	Town Hall Preparation	100.00	92.86	86.67	100.00	47.83	85.47
2. Identifying a Problem	Terminology Lecture (Perspective, Beliefs	100.00	100.00	86.67	90.00	50.00	85.33
4.3 Evaluating Content	Introduction to lesson 4.3	100.00	100.00	54.55		83.33	84.47
4.2 Evaluating Source	Taking a Crirical Stance Activity	73.33	100.00	76.67	87.50	78.57	83.21

Note. The criteria used to select the top 10 activities included activities delivered by at least four of the five interventions. It excluded Pretest and Posttest.

Table 16

Top 10 Activities with Lowest Student Engagement Scores – Researcher's Perspective

Lesson	Part	Engagement Score by Activity (Researcher's Perspective)					
		Kasey	Malena	Amy (51)	Amy (55)	Chris	M
6. CommunicatingFindinç	Writing a Thesis Statement	75.00		80.00	68.18	19.23	60.60
3. Locating Information	Intro	48.15		80.00	83.33	31.25	60.68
3. Locating Information	Lateral Reading	46.67	55.00	76.67	73.33	70.83	64.50
6. CommunicatingFindinç	Intro	78.57	91.67	60.00	70.00	30.77	66.20
3. Locating Information	Concepts Lecture & Activity	43.33	64.29	76.67	83.33	64.00	66.32
5. Synthesizing Informatic	Intro		86.36	94.44	55.00	29.63	66.36
4.2 Evaluating Source	Types of Organizations Lecture	67.86	57.14	83.33		57.14	66.37
4.2 Evaluating Source	Author's Expertise Lecture	57.89	100.00	70.00		41.67	67.39
4.3 Evaluating Content	Fact-checking with Snopes	100.00	100.00	20.00		50.00	67.50
4.3 Evaluating Content	Evaluating Content Credibility	90.00	94.74	34.48	58.33	60.87	67.68

Note. The criteria used to select the top 10 activities included activities delivered by at least four of the five interventions. It excluded Pretest and Posttest.

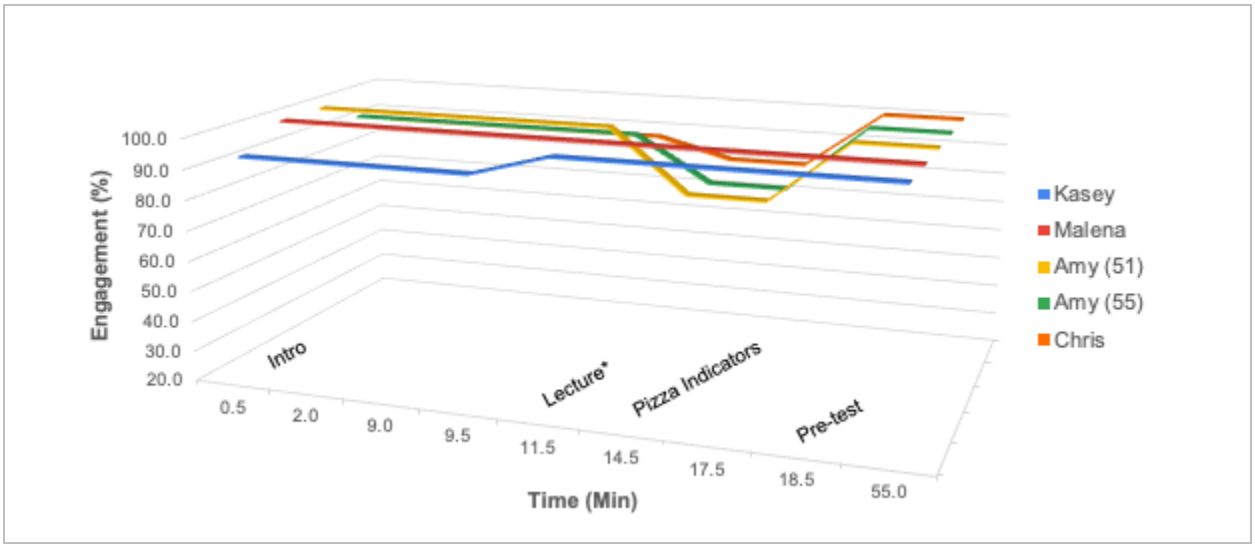
Student Engagement and Motivation by Activity (Researcher's Perspective.) The following section presents the students' engagement and motivation results by lesson and activity for all cases. The data was captured at this level of granularity (i.e., instructional activity)

only from the researcher's perspective, giving a deeper insight into the student's level of engagement by activity. This presentation perspective by Lesson and activity provides relevant insights for redesigning the CORE program for the next iteration of the intervention. *Lesson 1.1 – Introduction and Pretest* was the only one with a consistently high student engagement score across the five cases (Figure 18). The rest of the interventions had varied student engagement and motivation levels during the implementation. As one of the research observers stated: "The students in general move quickly to engaged to disengaged" (Aria). For example, during *Lesson 4.2 – Evaluating Source* (Figure 23), Kasey's and Malena's students demonstrated different ups and downs in their engagement scoring during the lesson evolution. At the same time, Amy's (55) score was relatively high and stable. On the other hand, Chris' class had a lower level of engagement, descending up to 28.6% during the Web of Understanding lecture.

Lesson 1.1 – Introduction and Pretest. This lesson had a very high level of engagement and motivation for all teachers (92.8%) from the researcher's perspective (Table 14). During the Pretest, all students seemed to be involved 100%. Figure 18 display the trending lines during development. The observers highlighted the Pizza Indicators activity as highly engaging.

Figure 18

Student Engagement and Motivation Lesson 1.1– Introduction and Pretest



* Evaluation concepts

Illustrative Quotes:

"ALL students were engaged talking about where to get the best pizza in Montreal … Even the silent child in the back chimed in" (Ellen)

"Amy (55) did an incredible job with the pizza indicators and the relationship with the credibility indicators. For example, she mentioned the "author" as being the cook of the pizza. If it is an Italian cook than it might predict a good pizza" (Aria)

"Most students were quite chatty, asking many questions about the project. They were excited by the project, and asking many questions about what event they could choose for the project. When the lesson 1.1 [Introduction and Pretest] content started, students seemed very energized and ready to go. A lot of students put their hands up as they wanted to share their answers" (Tristen).

"when he asks them [students] to focus, they do quiet down - although he has to do this many many times" (Tristen).

Case 1. According to Ellen, Kasey's students had a very high level of engagement (97.6%) during the first intervention lesson *1.1 – Intro*. Students participated in responding to the questions posed by the teacher. All students were engaged in the pizza activity. Nonetheless, during the pretest, students grumbled a little bit.

Case 2. Talya observed that Malena's class was 100% engaged during the first intervention lesson.

Case 3. According to Aria, Amy (51) had a high level of student engagement during *Lesson 1.1 – Intro* (93.1%). This group was characterized as quiet and disciplined, which showed during the lesson. Students had a high level of participation during the lecture activity, by writing their own definitions of credibility and relevance.

Case 4. Students in Amy's (55) class had a high engagement during *Lesson 1.1 – Intro* (90.2%), as said by Aria. During the Pizza Indicator activity students were very engaged. A highlight from this case is the proposed analogy Amy (55) gave between the 'author' and the 'cook' of the pizza.

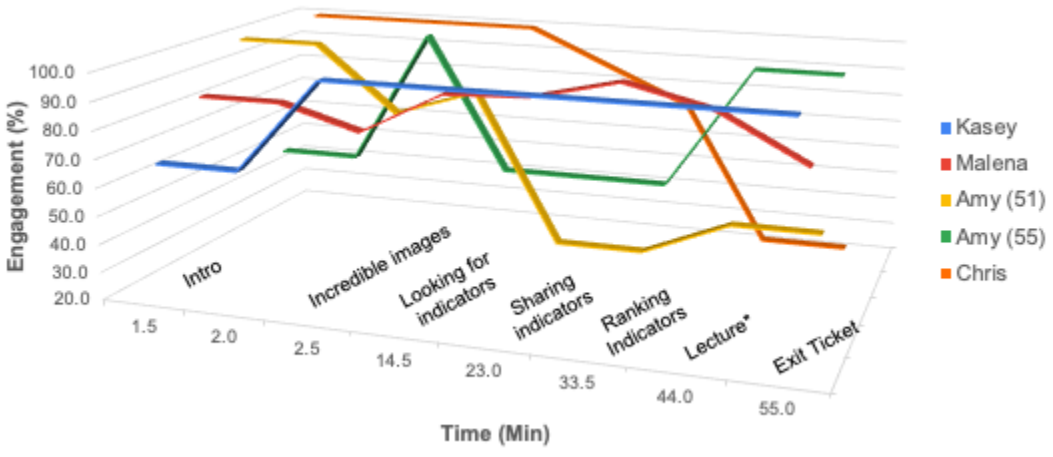
Case 5. Tristen observed that Chris' class had a high level of engagement and

motivation (83%). Students were interested in the project, asking many questions. There were a couple of students playing video games on their laptops, and some others seemed to be doing work for another class. The teacher needed to put a lot of effort into keeping the students engaged.

Lesson 1.2 – Overview. This lesson had on average a high level of engagement and motivation for all teachers (80.2%) from the researcher's perspective. However, the engagement by activity was not consistent across the cases (Figure 19). The Incredible Images activity had the highest scores (90.5%), followed by the Looking for Indicators activity (85.6%). The students lost interest by the end of the lesson during the Ranking Indicators activity (71.4%) and the Three Tier Framework lecture. The Exit Ticket, the last activity of the lesson, was scored with the lowest value of 68.2%. This result was highly affected by Chris' class scores since the school did not have internet, so the engagement level at the end of the class dropped to 23.1%.

Figure 19

Student Engagement and Motivation Lesson 1.2 – Overview



* The Three Tier Framework

Illustrative Quotes:

"The students were very excited about the content and had a great time doing the activities. You can see that they feel very safe being themselves, and the teacher encourages individual expression, as the students felt comfortable giving their opinions even if they conflicted with another student's" (Ellen).

"The students in general move quickly to engaged to disengaged. They multitask with their tablets and phones" (Aria)

"The school did not have wifi today, so the teacher had to improvise. His solution was to write the questions and activity prompts on the board, and have the students look at the articles and their workbook on their phones. Some students did not have cellular data, so they could not participate in this activity" (Tristen).

"As soon as a team leader was picked, the more engaged students seemed to take over and the less engaged students stood back and let the more engaged students take the lead… Once students started presenting, most of the class were engaged. After the first presentation, students seemed to be less engaged (students started talking to each other after their own group's presentation was done)" (Tristen).

Case 1. As perceived by Talya, students were highly engaged (94.5%) during *Lesson 1.2 – Overview* in Kasey's class.

Case 2. Ellen observed that Malena's class was highly engaged and motivated during *Lesson 1.2 – Overview* (87.8%). Students enjoyed the content and the activities. Students felt confident expressing their opinions, and they felt in a safe environment. During the group activities, students seemed to be highly engaged in discussing the topic (e.g., ranking indicators).

Case 3. As per Aria's observation, Amy's (51) class obtained a moderate engagement score (68.4%). Multitasking seems to be a characteristic of the group switching back and forth between tasks.

Case 4. Aria noted that the engagement score for Amy's (55) class was moderate (73.8%), driven mainly by the time of the class (4th period in the afternoon) and students feeling tired because they were preparing for a theatre play. Aria highlighted the Incredible Images activity as the most engaging activity of *Lesson 1.2 – Overview*. However, for the next activity of the lesson Part 1: Looking for Indicators, one student tagged the activity as a 'repetition' of the

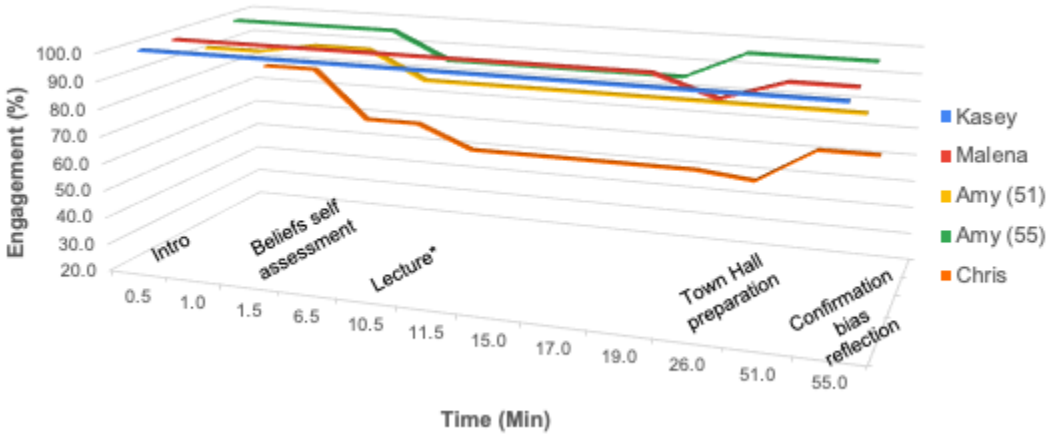
previous activity in *Lesson 1.1 – Introduction*.

Case 5. The engagement and motivation for this *Lesson 1.2 – Overview was moderate* (76.7%). It started high at the beginning of the lesson but went in decline as the lesson progressed, mainly driven by the lack of internet. Another important element noted by Tristen in the engagement aspect was choosing leaders for the Town hall activity; once the leader was selected, many students lost interest in the lesson.

Lesson 2 – Identifying a Problem. This lesson had, on average, a high level of engagement and motivation for all teachers (85.2%) from the researcher's perspective. The engagement by activity was relatively consistent across all cases (Figure 20). The initial activities of the class (i.e., presentation of the essential question and learning objectives) reported the highest levels of engagement (94.2%), followed by the Beliefs Self-Assessment activity (91.1%). Some observers emphasized the latter activity as one of the most interesting for students during this lesson.

Figure 20

Student Engagement and Motivation Lesson 2 – Identifying a Problem



* Perspective, beliefs, prior knowledge

Illustrative Quotes:

"Students seem to be quite restless, they shift quite a bit in their seats. Most students are not looking up at the teacher or looking at their peers when they are talking, they are looking down at their laptops. If they are not playing a game on their laptops, then they are looking ahead in their workbooks, scrolling through" (Tristen)

Case 1. Kasey's students were highly engaged (100%) throughout *Lesson 2 –Identifying a Problem*. Students were interested in the pandemic question and school closure, as observed by Talya. In addition, students were actively using the Student Workbook.

Case 2. Talya recorded a moderate engagement score (78.6%) during Malena's *Lesson 2 –Identifying a Problem*. She observed the teacher was keeping students engaged by asking questions about the terminology provided. The topic of confirmation bias was interesting to students. One important note during this lesson was the Covid topic 'fatigue.' Students showed interest in different social issues.

Case 3. As per Kelly's observation, Amy's (51) class obtained a high-level engagement and motivation score (90.8%) during *Lesson 2 –Identifying a Problem*, as per Kelly's observation. It is interesting the observer's note regarding the emphasis of the teacher on the utility of this intervention during CEGEP and the university will most probably use it to demonstrate the validity of sources for research projects.

Case 4. Kelly recorded a very high engagement and motivation score (97.5%) for Amy (55)' class during *Lesson 2 –Identifying a Problem*.

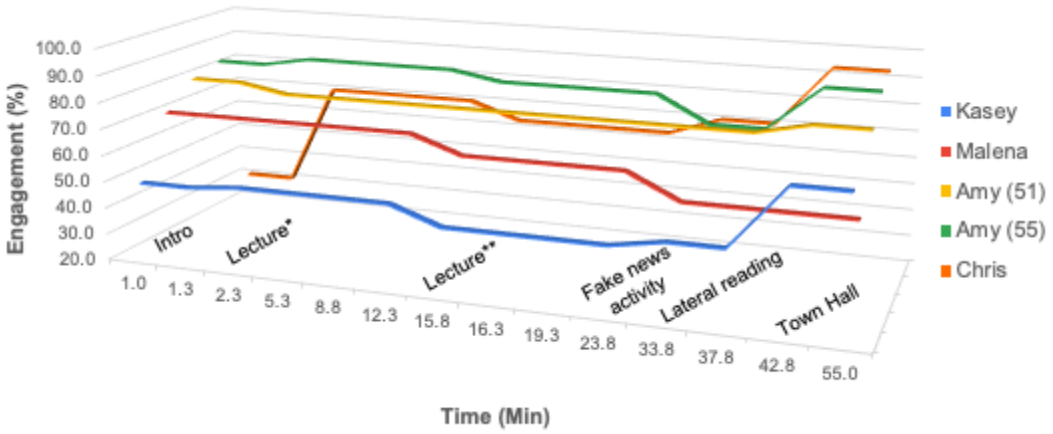
Case 5. Tristen observed moderate student engagement (59.2%) during Chris' *Lesson 2 –Identifying a Problem*. However, during the Beliefs Self-Assessment activity, students were particularly highly engaged. After that activity, students lost focus. Students were not totally engaged in the Town Hall preparation activity, partially explained by an official test of some sort. "The students seem very interested in this test and discuss it continually," said Tristen.

Lesson 3 – Locating Information. This lesson had, on average, a moderate level of

engagement and motivation for all teachers (67.1%) from the researcher's perspective. The engagement level by activity was inconsistent across all cases (Figure 21). During this lesson, student engagement and motivation were reported as the lowest scores of all the lessons, probably due to the complexity of the lesson concepts and the large component of direct instruction. However, we can see that all cases experience a rise in engagement and motivation towards the end of the lesson during the Town Hall activity, supported by the research observers in their notes. This result is consistent with the findings in the other two factors of adherence to content and quality of delivery, previously examined. An important highlight reported by one of the observers is that even though this lesson was at the end of the day, and the student seemed tired, they became engaged during the Town Hall activity.

Figure 21

Student Engagement and Motivation Lesson 3 – Locating Information



* Web, meta search engines, search engine, results page, algorithms, query syntax

** Filter bubbles, echo chambers, fake news

Illustrative Quotes:

"My impression was that the class had a very low level of engagement, getting them to do anything was extremely difficult for the teacher. It seemed like regular class activities were very difficult for students, let alone engaging with the course content" (Tristen).

"Students who are doing other things on their laptops seem to be distracting other students, these students are trying to look at what they are doing on their screens (Tristen).

"None of the ss [students] were engaged with the video [How fake news can spread activity]" (Ellen).

"The students were most engaged when they had an activity, they were doing such as the Town Hall, which was well-prepared for or answering the teacher's questions" (Ellen).

Case 1. Kasey's students registered a moderate level of engagement and motivation during *Lesson 3 – Locating Information* (51.6%). As this session was during the last period, students' energy seemed to be very low, mainly driven by the class at the end of the day. Students were disconnected from the class using their phones rather than focusing on the teacher. Kids who were playing games, navigating in Google Classroom, or working on homework for other classes on their laptops distracted other students who were trying to see what their peers were doing. During this lesson, the level of engagement escalated to its maximum level (70%) during the Town Hall activity.

Case 2. Ellen recorded a moderate level of engagement (56.4%) in Malena's *Lesson 3 – Locating Information*. None of the students showed interest in the video "How fake news can spread." Some observed students were totally disengaged by watching something else on their computers, cellphones, or just napping over the desk. "Maybe having them do more pair-work and small group activities would help."

Case 3. Kelly recorded a moderate engagement score (78.1%) for Amy's (51) class. The maximum level of engagement (81%) during this *Lesson 3 – Locating Information* was recorded for the Town Hall activity.

Case 4. Kelly recorded a high level of engagement (83.3%) during Amy (55)' class. The class participated in the Town Hall discussion with enthusiasm. The teacher engaged the students with a voting system by which they could assess their peers and elect the "most"

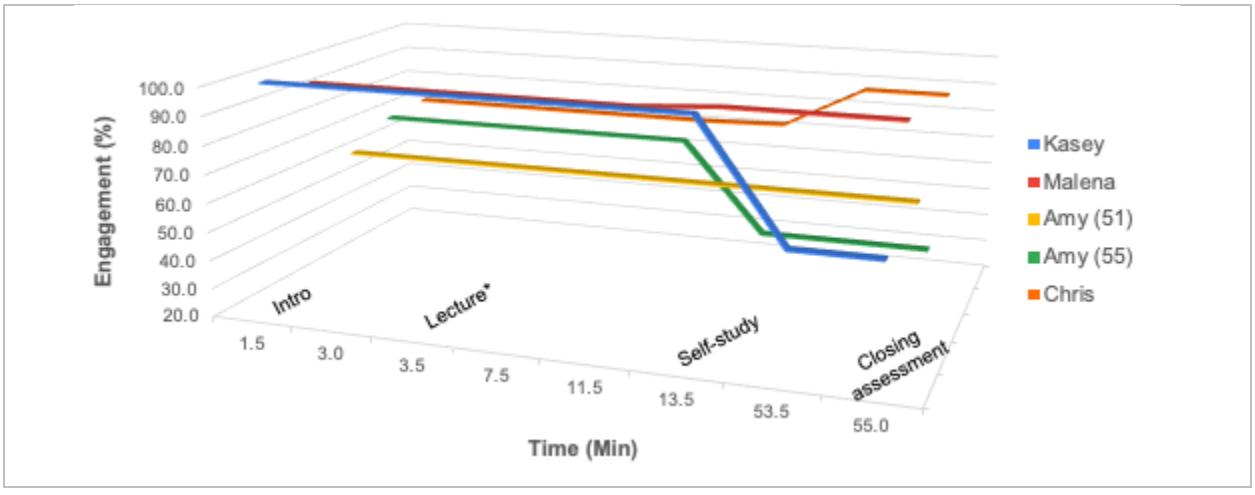
credible role judged by the presented arguments.

Case 5. Ellen recorded a moderate level of engagement (66%) during Chris's presentation of *Lesson 3 – Location Information*. The observer recorded the Town Hall activity as the highest engagement score (93.3%). "The students were most engaged when they had an activity they were doing such as the Town Hall, which was well-prepared for or answering the teacher's questions," Ellen noted.

Lesson 4.1 – Evaluating Context. This lesson had, on average, a moderate level of engagement and motivation for all teachers (75.1%) from the researcher's perspective. The engagement level by activity was inconsistent across all cases (Figure 22). This class was a self-study lesson, and the evidence suggests students did not seem excited about this instructional approach. This finding was corroborated by the teacher's perspective, where some of them perceived the students were not motivated by the lesson as it did not have interactions from their perspective. Moreover, many students voiced in the Student Survey that one of the things they liked the most was the group activities which this specific lesson did not offer (see more detailed results on students' engagement and motivation in the following section, *Student's Perspective*.)

Figure 22

Student Engagement and Motivation Lesson 4.1 – Evaluating Context



* The Three-tier framework lecture

Illustrative Quotes:

"I believe that student engagement is measured as high because they have questions on how to navigate the lesson rather than the content of the lesson" (Talya)

"Students were not engaged in the self-study. Some others did not have earphones. The class was disrupted by another teacher discussing a problem, and the student became 'upset' and lost focus on the lesson" (Aria).

"As per usual, there are several students playing games on their laptops (I can see 3 students consistently playing games, 2 students switch between a game and the activity). When the activity instructions are presented, students become more engaged (as shown in round 3) and ask the teacher a lot of questions regarding how they are supposed to complete the activity" (Triten).

Case1. During *Lesson 4.1 – Evaluating Context* (self-study), Kasey's students did not show enthusiasm, as recorded by Talya. On average, the level of engagement was high (86.7%), starting wight a very high number of engaged students (100%). However, as it progressed, it dropped to a low level (60%) by the end of the lesson.

Case 2. Malena's class had a very high level of engagement (95.4%), as Talya recorded. She observed that this could be explained more by the "questions on how to navigate the lesson rather than the content of the lesson."

Case 3. Amy's (51) class had a moderate level of engagement (62.5%) during *Lesson 4.1 Evaluating Context*. It is important to note that half of this class was used to finish the Town Hall activity from the previous lesson; thus there are not sufficient observation notes related to lesson 4.1.

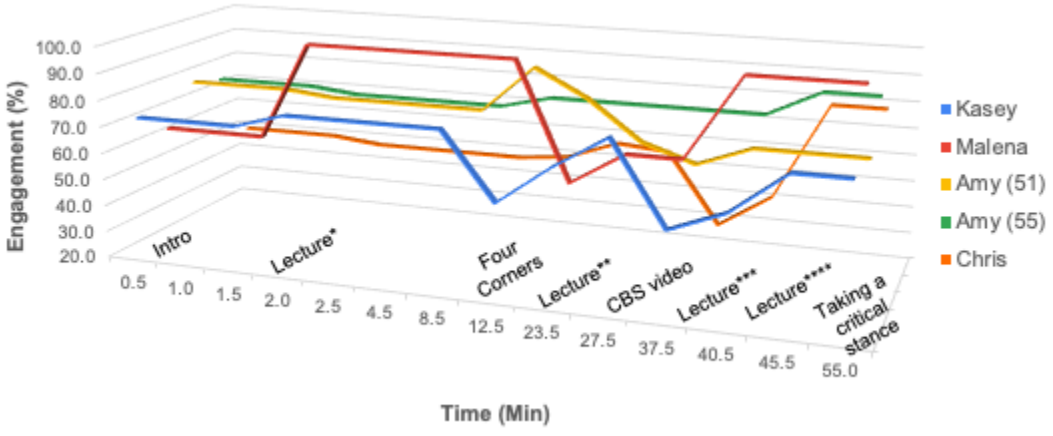
Case 4. Aria recorded a moderate level of engagement (53.8%) during Amy (55)'s class. The class was generally somewhat noisy; students were not quietly following the lesson. In general, students were not engaged with the self-study lesson. In addition, a teacher from another class came to discuss an external issue with the students, causing discomfort and disengagement.

Case 5. Chris's students were moderately engaged (77%), as perceived by Tristen. Student engagement goes on and off, depending on the activity. Tristen also observed that this class was smaller due to missing students. making it easier to control and thus increasing engagement "as fewer students were calling out and getting distracted."

Lesson 4.2 – Evaluating Source. This lesson had, on average, a moderate level of engagement and motivation for all teachers (71.1%) from the researcher's perspective. The lesson attained the second-lowest score across all lessons for all cases. The engagement and motivation level by activity was very inconsistent across all cases. In Figure 23, we can appreciate the student's engagement and motivation scores in a kind of 'rollercoaster' manner, with ups and downs. By the time this lesson was conducted, students started to lose interest in the program, as corroborated by the teacher's perspective findings. The observers noted that some teachers used other facilitation strategies to keep the student engaged, such as randomly selecting students to participate, physically making the Four Corner activity instead of using the Jamboard or adapting the activities.

Figure 23

Student Engagement and Motivation Lesson 4.2 – Evaluating Source



* CORE concepts (evaluation, credibility, relevancy)

** Types of Organizations

*** Web of Understanding

**** Author's Expertise

Illustrative Quotes:

"When the teacher asked students to present their ideas, she did not have to wait for a response or ask someone to present, students were ready to give their response right away" (Tristen).

"Engagement score is high because this is a student activity" (Talya).

"Students were very engaged in the activity [Four Corners activity.] They supported their decision very well around the primary purpose definition" (Aria).

"Some of the computer games are distracting even for me (bright colours and things moving around the screen encourage my eyes to look at the student's screen more). So, not only is the student distracting themself from the lesson, but they are also distracting everyone around them" (Tristen).

Case 1. Tristen recorded a moderate engagement and motivation level (66.6%) during *Lesson 4.2 – Evaluating Source* in Kasey's class. During this lesson, students seemed energized and ready to discuss in Kasey's class. At this point of the program, Kasey expressed students started to get 'fatigued' with the intervention.

Case 2. Talya logged a high level of engagement (84.3%) for *Lesson 4.2 – Evaluating the Source* in Malena's class. "Engagement score is high because this is a student activity," observed Talya. Another teacher's strategy to increase student engagement was randomly selecting students to participate in the discussion.

Case 3. Amy's (51) class had a moderate level of engagement and motivation (76.1%), as observed by Aria. Students were highly engaged in the Four Corners activity and they proposed good arguments to support their corner election, reflecting the 'primary purpose' definition.

Case 4. Amy's (55) level of engagement and motivation was moderate (77.7%) as noted by Aria. The teacher used several strategies to involve students, such as adapting the activities to be done in groups.

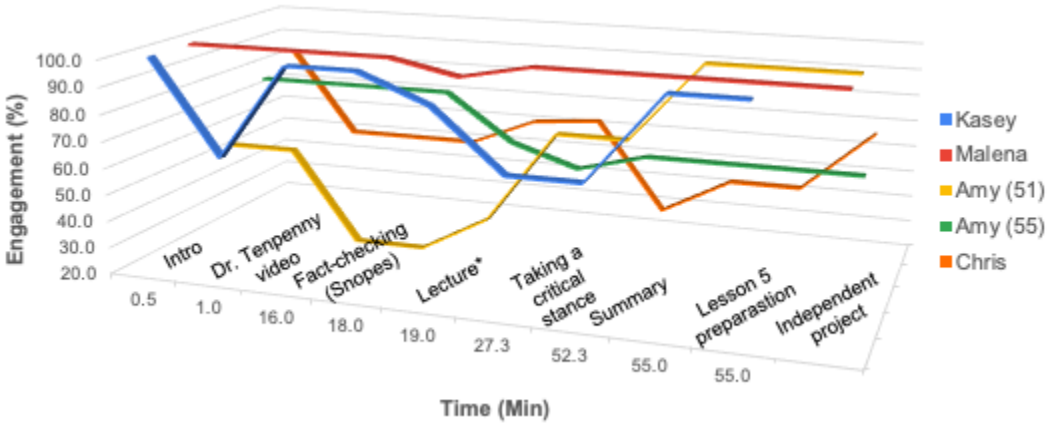
Case 5. Tristen observed a moderate engagement with a downward trend (51.2%) in

Chris' class. It could be explained by the lesson in the last period, "so engagement seems to be naturally low," noted Tristen. As usual, some students were playing video games on their laptops.

Lesson 4.3 – Evaluating Content. This lesson had, on average, a moderate level of engagement and motivation for all teachers (72.9%) from the researcher's perspective. The engagement level by activity was inconsistent across all cases (Figure 24). Likewise, for *Lesson 4.2 – Evaluating Source*, the engagement and motivation scores for this lesson seem to have a rollercoaster effect, where the group student activities reached the highest points and the lectures the lowest. As recorded for other classes and corroborated in previous findings from the teacher's perspective, external classroom factors affected student engagement and motivation. For this specific lesson, a physics exam following the CORE class and a student talent show spectacle the previous two nights are shown, as evidenced by the classroom context's impacts on the studied variable.

Figure 24

Student Engagement and Motivation Lesson 4.3: Evaluating Content



* Content credibility

Illustrative Quotes:

"Group activities appear to make students active during the lesson time" (Talya).

"I had the opportunity to discuss with Amy about engagement of students in a day like today that they have an exam next period. She was able to manage the class and engage the students to certain extent. We recognize that nowadays students have the ability of multitasking. When you observe them, many times they do not look engaged, but when I was checking the Google form responses, I saw they were doing the task" (Aria).

"One student even commented at the start of the lesson when they saw the first slide – 'I was doing fine before I saw this, are we really still doing this'. Compared to their usual level of engagement, I would say the students were slightly more engaged (that is why I gave them an engagement score of medium)" (Tristen).

Case 1. The overall lesson engagement was high (90.3%). Specific signals might explain the high score for student engagement, such as the videos, sharing the results and the group activities. During *Lesson 4.3 – Evaluating Content*, group activities seem to make students more active.

Case 2. Student engagement and motivation scored for Malena's class in *Lesson 4.3 Evaluating Content* was very high (99.4%). According to Talya, this outstanding rating might be explained by students being required to fill in documents, group activities, sharing the results to the classroom, and the specific topic of fact-checking tools such as Snopes.

Case 3. Aria reported moderate (58.3%) student engagement and motivation scores during Amy's (51) *Lesson 4.3 Evaluating Content*. The main reason for this engagement score might be the students' physics exams in the next period. Aria had the opportunity to discuss with the teacher about engagement during the intervention, and she highlighted the topic of multitasking as a characteristic of current high school students.

Case 4. Aria logged Student engagement and motivation as moderate (60.5%) during Amy's (55) *Lesson 4.3 – Evaluating Content*. Students felt very tired as they had participated in a talent show the previous night.

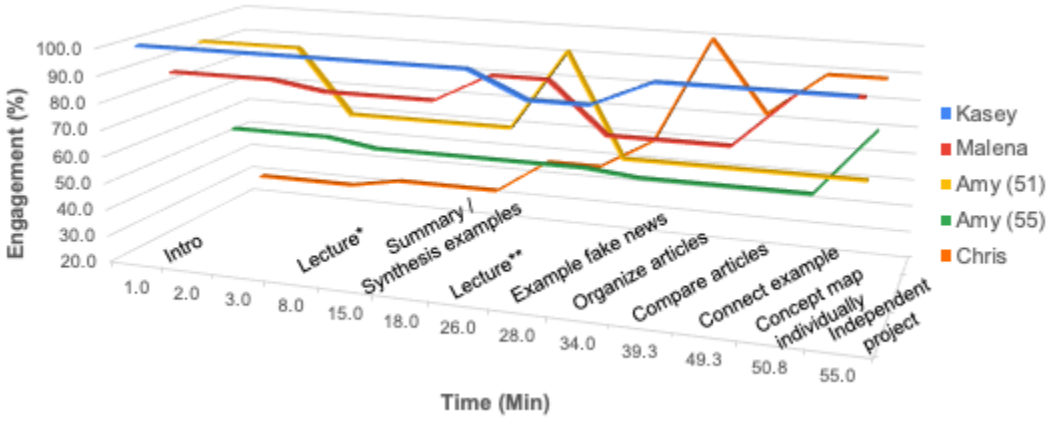
Case 5. Tristen recorded a moderate rate of student engagement and motivation (55.7%) in Chris' class. Students seemed calm and focused, but they started feeling fatigued

with the overall intervention.

Lesson 5 – Synthesizing Information. This lesson had, on average, a moderate level of engagement and motivation for all teachers (76.1%) from the researcher's perspective. The engagement level by activity was inconsistent across all cases (Figure 25). During this lesson, most of the teachers skipped several activities for different reasons, including but not limited to, lack of time, students' prior knowledge, or accommodation of the teacher due to students' tiredness from a sports event the previous night. Student displayed the highest engagement during the Create Concept Map activity. A note to highlight is that Kasey graded this activity, obtaining a perfect score (100%) on student engagement and motivation, emerging as a potential factor that moderates such results.

Figure 25

Student Engagement and Motivation Lesson 5 – Synthesizing Information



* Summary vs. Synthesis

** Concept map

Illustrative Quotes:

"When students are asked to work together, the engagement score is often higher, there is more class discussion and students are more interested in the lesson" (Talya)

"The teacher and the students wanted to get through these lessons so that they could move on to assignments that counted towards more points" (Ellen).

"ALL of the students are playing video games during round 1. Teacher brought them back to attention with calling out by name, CCQs [content checking questions], expansion questions and open class question. Students are multi-tasking between games and answering the teacher's questions during the activity" (Ellen).

Case 1. The overall student engagement score for *Lesson 5 – Synthesizing Information* was very high (98.5%) for Kasey's classroom, as reported by Talya. When asked to work together, students were more interested in the lesson, "the engagement score is often higher." Another moment of higher engagement was when the students completed their own concept maps for their projects. In Kasey's instance, it was collected and graded, which might explain the high score for engagement.

Case 2. Ellen reported a moderate student engagement score (87%) for *Lesson 5 – Synthesizing Information*. Malena blended lessons five and six because the previous class was missed due to a snow day and two days before spring break. Both students and teachers wanted to complete the program quickly to move to other regular activities that counted for grading. All the students were engaged during their urgency of completing everything on time.

Case 3. Aria reported Amy's (51) engagement score during *Lesson 5 – Synthesizing Information* as moderate (75.4%). Students were tired, and many were absent since there was a game. The teacher went quickly over the material as she planned to repeat the session the following week.

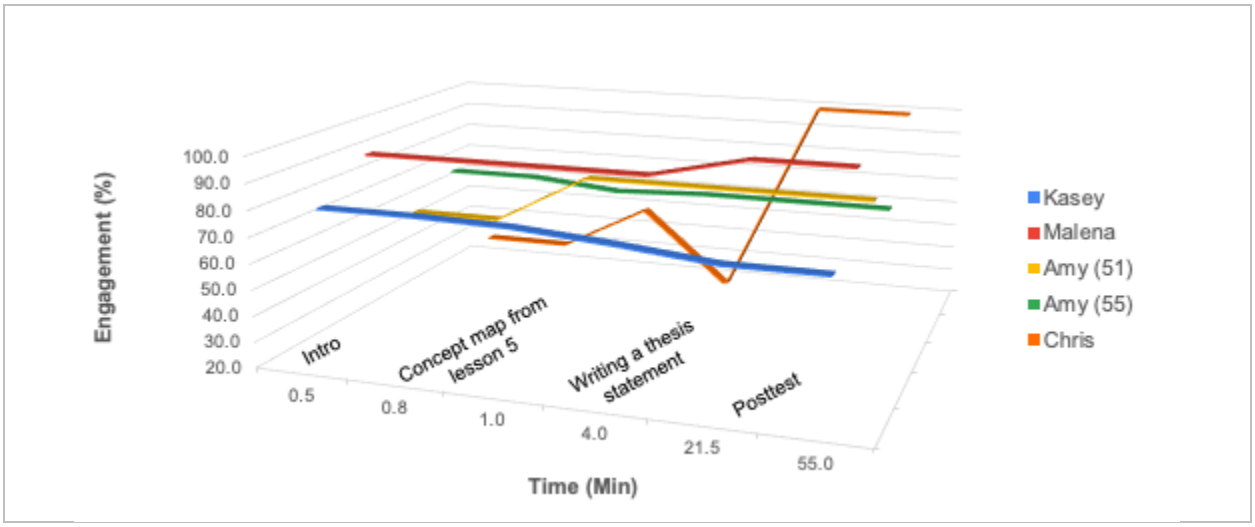
Case 4. Student engagement during the lesson was moderate (58%) for Amy's (55) classroom, as reported by Aria. The activity that engaged most of the students was the evaluation of the sources for the independent project; this might be partially explained by the fact that it was going to be graded.

Case 5. Ellen recorded a moderate student engagement score (61.7%) during Chris' *Lesson 5 – Synthetizing Information.* This class had learned concept mapping earlier in the year, "it might account for their lack of engagement in the CORE classes," noted Ellen. During the introductory activity, students had a very low engagement rating (29.6%) as they were playing video games. All the students were engaged while working with peers and doing their work. This behaviour was evident during Activity 1 [Organize Articles] and Activity 2 [Compare Articles], where the student engagement score was 100%.

Lesson 6 – Communicating Information. This lesson had, on average, a moderate level of engagement and motivation for all teachers (72.6%) from the researcher's perspective. The engagement level by activity was inconsistent across all cases (Figure 26). The student reached the highest engagement score on responding to the Posttest, especially in Chris' class (100%), most likely explained by the fact that the teacher explicitly communicated to the students his high expectations on behaviours and achievement (see the quality of delivery section). Another note highlighted by one of the research observers that might impact students' engagement scores was the proximity of the spring break (two days), where students were engaged and rushed to finish all their assignments before the break.

Figure 26

Student Engagement and Motivation Lesson 6 – Communicating Information



Illustrative Quotes:

"Students were behind in their homework. They were supposed to have come to class with their thesis statement and topic ready. The majority of the class had not done this which frustrated the teacher. She had to spend extra time re-explaining how they should choose their topics and write their thesis statements" (Tristen).

"It was also 2 days before the spring break, so everyone was in a rush to complete everything that they needed to complete" (Ellen).

"They all seemed very concerned about marks and are obviously strategic about using attention and time. It is possible that they do not see a value of placing large amounts of attentional resources to something that is interesting but that doesn't provide them with marks" (Ellen).

Case1. Kasey's class achieved a moderate student engagement score (75.7%) as stated by Tristen. The teacher seemed to have difficulty facilitating *Lesson 6 – Communicating Information*, because students did not prepare for the class and were behind their homework.

Case 2. Ellen rated student engagement and motivation as high (95.8%) for Malena's class. The class was mainly engaged in the Posttest and completing any extra homework they had not finished. "It was also 2 days before the spring break, so everyone was in a rush to complete everything that they needed to complete," noted Ellen.

Case 3. Aria recorded a moderate student engagement and motivation score (73.3%) during Amy's (51) class for *Lesson 6 – Communicating Findings class*. Students were engaged in writing the Posttest.

Case 4. A moderate student engagement and motivation score (68.3%) was recorded by Aria during Amy's (55) class for the lesson.

Case 5. According to Ellen, students were disengaged during *Lesson 6 – Communicating Finding* (50%) in Chris' class. Ellen's interpretation of disinterested behaviour is that students may invest time and energy in graded activities. All the students were focused on the posttest.

Participant Responsiveness – Student's Perspective

The following section presents the results of participant responsiveness in two parts. The

first part presents the quantitative results we obtained after running a one-way analysis of variance (ANOVA) to determine whether there were any statistically significant differences between the student engagement and motivation means of the four independent groups[5] (i.e., classrooms). The continuous dependent variable was student engagement and motivation scores, collected from students and measured on a scale from 1 to 100. The independent variable was the classroom (i.e., Amy [51], Chris, Kasey and Malena). In the second part, we present the qualitative results obtained from three open-ended questions. We ask students to mention the things they liked the most from the CORE program, the things they liked the least, and other topics that would interest them. Note that the student engagement and motivation from the student perspective were measured at the intervention level rather than at the lesson level, like the teacher's perspective, or the activity level, like the research perspective. This survey was conducted only once towards the end of the CORE program, and 95 students completed it.

Quantitative Results. Students had a moderate engagement and motivation level (69.6%) from the student's perspective (Table 17). Only one teacher, Chris, achieved a high-level score (84.6%), reporting a smaller standard deviation value as well (*SD* = 10.5). A further one-way ANOVA (Figure 27) was conducted to determine if the student engagement and motivation score was different for a classroom with different teachers[6]. Students were classified into four groups: Amy (*n* = 13), Chris (*n* = 6), Kasey (*n* = 43) and Malena (*n* = 33). There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by visual inspection of Normal Q-Q Plots; and variances were homogeneous, as assessed by Levene's test of homogeneity of variances (*p* = .719). Data is presented as mean ± standard deviation. Student engagement and motivations were statistically significantly different between

[5] Only students from Group 51 in Amy's classes responded to the Student Engagement and Motivation Survey.
[6] See Appendix C for detailed supporting documentation of the ANOVA test.

different classrooms, F (3, 91) = 3.320, p = .023, ω^2 = 0.068. Student engagement score increased from Kasey's (M = 67.74, SD = 11.70) to Amy's (M = 68.95, SD = 18.86), Malena's (M = 69.42, SD = 12.58) and Chris' (M = 84.55, SD = 10.46) classrooms, in that order. Tukey post hoc analysis revealed that the mean increase from Malena to Chris (15.12, 95% CI [.89, 29.36]) was statistically significant (p = .033), as well as the increase from Kasey to Chris (16.81, 95% CI [2.83, 30.78], p = 0.12). However, no other group difference was statistically significant.

Table 17

Student Engagement Scores – Student's Perspective

Teacher	N	M	SD	Engagement Level by Teacher
Amy[a]	13	68.95	13.86	Moderate
Chris	6	84.55	10.46	High
Kasey	43	67.74	11.70	Moderate
Malena	33	69.42	12.58	Moderate
Total	95	69.55	12.70	Moderate

[a] Scores for group 51. Data was not available for group 55.

Figure 27

One Way ANOVA – Student Engagement and Motivation by Teacher (classroom)

Total Student Engagement Score %

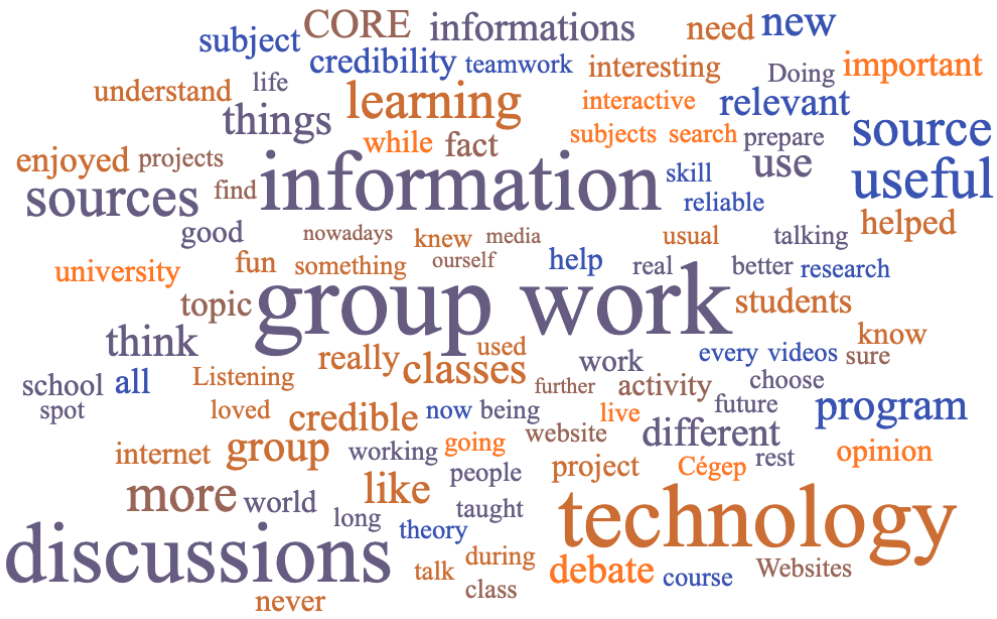
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1495.599	3	498.533	3.320	.023
Within Groups	13664.523	91	150.160		
Total	15160.122	94			

Qualitative Results. To the question, describe what you liked the most about the CORE program, the students responded that they liked the most group activities and discussions (Figure 28). Working in groups allowed them to share their ideas and perspectives, and it was valuable to collaborate with classmates and hear different viewpoints on the topics they were studying. Additionally, the program taught them how to identify credible sources and evaluate

the reliability of information, which is a useful skill for their higher education pursuits and life.

Figure 28

Things Students Liked the Most – Student Engagement and Motivation



> "I like how it taught me how to identify whether an article is credible or not. I also liked how there were many group works." (Student 33)
>
> "I really liked the working in group aspect of this program because I feel like you learn more when you share your ideas with people. I like the different types of activities we did to better understand what we were learning. " (Student 17)
>
> "I liked the part where we discovered how to debunk fake news and how to spot it within articles." (Student 37)
>
> "We learned alot of good information about this topic and it was very good to learn as this will help us in the future. I also enjoyed how much we got to learn as a group and do some of the tasks together." (Student 129)
>
> "It helped me to understand the topic better and really showed me the truth behind these sites and how they work." (Student 124)

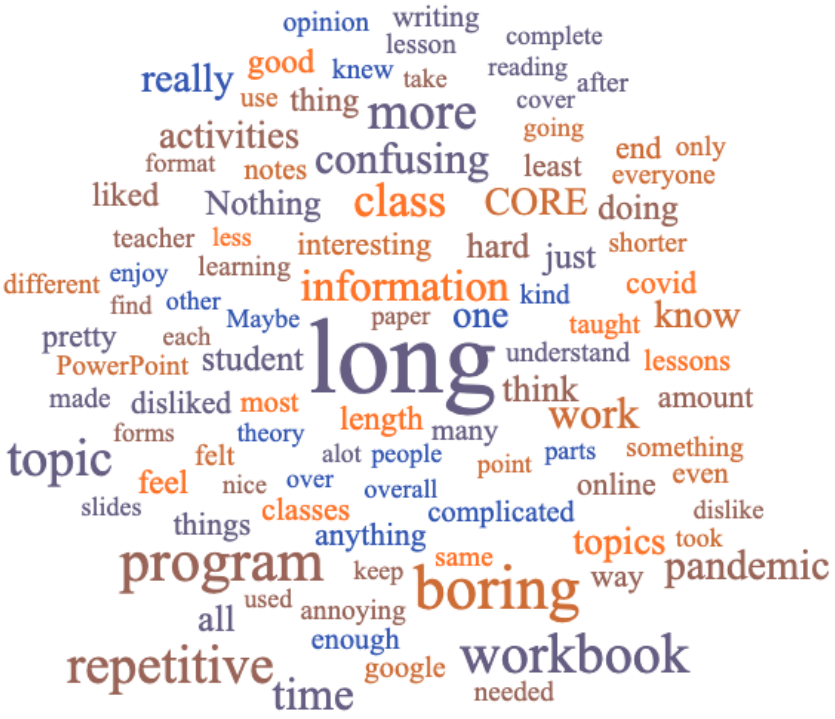
To the question, describe what you liked the least about the CORE program, the

students included the following: the length of the program, repetitive topics, difficulties with the online Student Workbook, unclear lessons, uninteresting pandemic topic, excessive amount of work, confusing and boring workbook, and the feeling of already knowing most of the information (Figure 29). Some students also mentioned technical issues, uninteresting surveys, and the need for more variety in topics.

> “I didn’t like that it took to long to finish the program.” (Student 33)
>
> “The pandemic topic ” (Student 4)
>
> “What I liked least about this program was that sometimes the topics were repetitive, like the, should schools open? I also feel like the student workbook can be improved because a lot of the graphs we can’t even edit, so you have to end.” (Student 57)
>
> “This program , though useful, was incredibly long and painful to sit through. Sometimes, so much information come flying at you that you don’t have time to internalize what you just heard.” (Student 25)
>
> “I did not like the online student work book I would have preferred a paper workbook because the online one was very complicated and hard to follow.” (Student 37)

Figure 29

Things Students Liked the Least – Student Engagement and Motivation



To the question, would you please recommend any other topic of interest for this CORE program, the topics that reflect the students' interests include politics, social justice, science, discrimination, sports, education, climate change, and mental health (Figure 30). Some students felt that the program focused too much on COVID-19, a topic they were already familiar with and tired of. Overall, most responses expressed positive aspects of the CORE program; some students had neutral or mixed opinions.

Figure 30

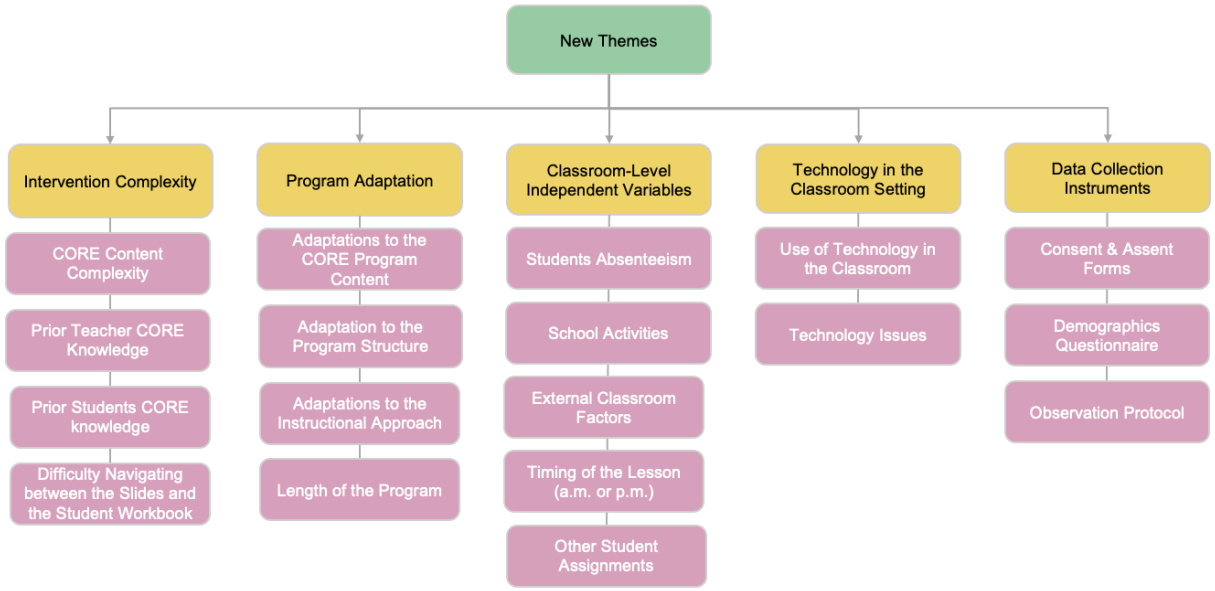
Students Topics of interest for the CORE program

Emerging Patterns and Themes

While working with the responses to the open-ended survey collected from teachers, researchers, and students, we noted recurring patterns or themes. The patterns constructed from the mixed methods data resulted in 18 categories that were synthesized into five major themes: Intervention Complexity, Program Adaptation, Classroom-Level Independent Variables, Technology in the Classroom Setting, and Data Collection Instruments (Figure 31). The following sections provide a high-level description and findings associated with the new themes. The following sections provide a high-level description and findings associated with the new themes.

Figure 31

Qualitative Analysis – Emerging Patterns and Themes



Intervention complexity

Intervention complexity is a potential moderator of program implementation fidelity as defined in the Conceptual Framework for Implementing Fidelity (Carroll et al., 2007). Although this variable was out of scope for this study, the theme emerged during the mixed method analysis. Detailed or specific interventions are more likely to be implemented with high fidelity than vague ones (Carroll et al., 2007). Furthermore, there is evidence that it is easier to achieve higher levels of fidelity with simpler interventions (Dusenbury et al., 2003, as cited in Carroll et al., 2007).

CORE content Complexity. Teaching the CORE program to high school students in Quebec resulted in a more complex task than initially anticipated. This intervention was a comprehensive and structured nine-lesson program with various degrees of complexity. The teachers mentioned *Lesson 3 – Locating Information* as one of the most complex and challenging lessons. The terms were hard to explain and required extra preparation before class and more explanation while teaching the lesson. Furthermore, some teachers mentioned they

adapted the presentation slides to include explanations and details. Some other CORE program components emerged as easy to deliver by the teacher, such as *Lesson 5 – Synthesizing Informatio*n and *Lesson 6 – Communicating Findings*, where the teacher found themselves comfortable and proficient in explaining, mainly because they had taught those topics previously.

Prior Teacher CORE Knowledge. Some teachers seemed more familiar with the topic than others and comfortable conducting the intervention. This depth of understanding allowed them to teach the complex concepts of the CORE program with more confidence. For those teachers less familiar with the complex concepts, the evidence suggests they dedicated a significant amount of time to preparation. Although they mentioned being nervous, there are no notes about the lack of confidence in delivering the lesson recorded by the observers.

Prior Students CORE knowledge. The prior knowledge and skills of students also play a role. Per the teacher's communication, some students were previously exposed to evaluating online sources, so teaching them the CORE advanced knowledge might have been less complex. However, teachers needed more time to review the intervention content for students with less exposure to the topic to gradually build their understanding.

Difficulty navigating between the Slides and the Student Workbook. One complexity was added to the CORE program, as mentioned by teachers and students, and it was the difficulty to navigate and toggle between the Presentation slides and the Student Workbook. Students found the workbook confusing and boring. Teachers had a challenging time when the slides did not match the workbook, such as the preparation for the lesson 5 activity covered in *Lesson 4.3 – Evaluating Content*. Students had difficulty adding text to some workbook parts (e.g., *Lesson 1.2. – Overview*, p. 10).

Program Adaptation

Adaptation is an essential component of putting intervention programs into action (Carvalho et al., 2013). Building this understanding into the programs might be challenging but

essential for maximizing its effectiveness in diverse contexts. According to O'Donnell (2008), program developers o researchers need to understand, for future implementations, which components are essential and require the highest level of fidelity and which ones can be adapted, modified, or removed. Programs must account for teachers adapting the intervention to the population's cultural, linguistic, and contextual factors while pursuing fidelity.

Adaptations to the CORE Program Content. Overall, the teacher delivered the program content as designed. Most of the time, teachers skipped (i.e., removed) content due to timing issues and changes derived from the program structure adaptation (see section below). Minor adaptations were recorded, such as adding text and details to the presentation slides, changing covid related examples, and adding new examples more relevant to the class context.

Adaptation to the Program Structure. Malena made one major adaptation to the program structure with a significant impact on the results (as seen in the adherence and quality of delivery scores). She blended *Lesson 5 – Synthesizing Information* and *Lesson 6 – Communicating Findings* into one lesson because students had prior knowledge of both topics. This adaptation caused a decrease in the adherence to content score, a lack of time to deliver the full content of the program appropriately, and, more importantly, diminished the time students dedicated to the Posttest. Kasey and Malena made another example of adapting to the program structure by substituting the Independent Project (proposed by the CORE program) and the Persuasive Paper (regular assignment of teachers' English class). While the adherence score decreased, this could be considered a fair adaptation of the program to the class context, given that this was suggested as an optional activity in the program. Minor adaptations were made by Amy (51) before delivering the intervention the second time (Group 55). As she learned from the first time, she accommodated the flow of the lessons, the material, and the timing. During *Lesson 2 – Identifying a Problem,* Kasey modified the lesson flow by initiating the Town hall activity before teaching the concepts, allowing her to expand on how they found the sources. As per her judgement, this was a better approach.

Adaptations to the Instructional Approach. Overall, teachers followed the instructional design proposed by the CORE program team, a hybrid direct instruction approach with cooperative learning. However, it was voiced loud by the students' preference towards group activities, discussions, and sharing ideas and opinions.

Length of the Program. One clear emerging theme from teachers' and students' perspectives was the length of the intervention. From their point of view, it was overly lengthy, causing fatigue, boredom, and loss of interest on the student side and time constraints for teachers who had other academic responsibilities to fulfill in their regular classes. Consequently, many of the adaptations and deletions to the content (i.e., skipping) made by teachers were in response to the time constraints experienced by the teachers, plus the extra effort in managing less engaged and motivated students.

Classroom-Level Independent Variables

Several factors within the classroom environment can significantly influence the implementation fidelity of an intervention (Combs et al., 2022). These authors explored the association between implementation fidelity and malleable variables within the classrooms that could be targeted to optimize resources as the school-based intervention disseminated across real classroom settings. Their findings suggest that curriculum modifications, students' misbehaviour, and shortage of time to implement the program were factors most associated with lower levels of implementation fidelity. Class size and access to program materials were less predictive. The following themes emerged during the CORE intervention and identified specific factors that might be associated with fidelity implementation.

Students Absenteeism. During the intervention, students were frequently absent for various reasons, including ice fishing day, covid, sports games, Passover holiday celebrations, snow days or personal motives. Absenteeism was a critical factor during certain activities, such as the Pretest during the first lesson, the Town Hall activity, and the Posttest in the last lesson. Absence for the Town Hall was specifically challenging for Case 1 – Malena since groups were

assigned during *Lesson 2 – Identifying a Problem*. However, half of the class was absent during *Lesson 3 – Locating Information* when the Town Hall took place. Malena had to redistribute the teams and needed to allow time to prepare again. This absenteeism impacted the program's overall delivery, which was already impacted by time constraints. Students who were absent during the previous lesson struggled as well with this Town Hall activity since they did not belong to any group and had to rely on their peers who were there.

School Activities. Several school activities ran parallel with the CORE program intervention, such as pyjama day, talent show, and sports games. Although these activities positively affect the student's integral development, they entailed some consequences for properly developing the CORE intervention. For example, during *Lesson 4.2 – Evaluating Source* and *Lesson 4.3 – Evaluating Content*, Amy (55)' students conducted a talent show for two consecutive nights. Students became more focused on the event and diverted attention from the lesson objectives. Furthermore, after the event, the students were exhausted, leading the teacher to adapt the delivery of the lesson in a way not to get the students tired, skipping some activities, changing activities approach (e.g., individual activity changed to group activities, interactive activity change to direct instruction), and even postponing part of the lesson for the next class.

External Classroom Factors. During the CORE program intervention, several external classroom factors impacted the learning environment for the students, such as a car alarm going off, another teacher interrupting the class to scold the students, beautiful spring weather outside, and lack of WIFI in the school. These external classroom events not only disrupted the regular flow of the lesson but posed a time management challenge for the teachers, who had already been constrained by the length of the intervention. Furthermore, in Chris' case, the lack of WIFI during the entire *Lesson 1.2 – Overview* forced him to improvise by asking the students to look at the articles and their workbooks on their phones. Some students did not have cellular data, so they could not participate in this activity.

Timing of the Lesson (a.m. or p.m.). From the total instances of the intervention lessons (N = 45), 49% were delivered in the morning and 51% during the afternoon (Table 18). Analyzing the responses from the Teacher's Journal and the Observation Protocol made it evident that the timing of the lesson, morning or afternoon, had implications for the intervention program. Both researchers and teachers concurred that students were more attentive and alert in the morning than in the afternoon. The observers frequently mentioned the higher levels of student energy during the morning, tiredness, and fatigue during the afternoon, accompanied with corresponding comments of higher levels of student engagement and motivation during the early class periods and less towards the end of the day.

Table 18

Timing of the Intervention

	Time				
Lesson	Kasey	Malena	Amy (51)	Amy (55)	Chris
1.1 Intro & Pretest	A.M.	A.M.	A.M.	P.M.	P.M.
1.2 Overview	P.M.	P.M.	A.M.	P.M.	P.M.
2. Identifying a Problem	P.M.	P.M.	P.M.	A.M.	A.M.
3. Locating Information	P.M.	P.M.	P.M.	A.M.	A.M.
4.1 Evaluating Context	A.M.	A.M.	A.M.	P.M.	A.M.
4.2 Evaluating Source	A.M.	A.M.	A.M.	P.M.	P.M.
4.3 Evaluating Content	A.M.	P.M.	P.M.	P.M.	P.M.
5. Synthesizing Information	P.M.	A.M.	P.M.	A.M.	A.M.
6. Communictaing Findings	P.M.	P.M.	A.M.	A.M.	A.M.

Other Student Assignments. Teachers detected with some frequency students working on assignments and homework for other classes. For example, Amy (51) mentioned that during *Lesson 6 – Communicating Finding,* some of her students focused more on the Ministry's French exam the next day than the CORE lesson. The researchers also noted a few times students working on their other class assignments, such as math homework and physics exam. When the students simultaneously worked on other assignments, their attention became divided, and they were not fully engaged with the intervention activities. One key point highlighted by Ellen, one of the observers, is the impact of grading on student engagement, "They all seemed very concerned about marks and are strategic about using attention and time.

It is possible that they do not see the value in placing large amounts of attentional resources on something interesting but does not provide them with marks."

Technology in the Classroom Setting

The CORE program was designed with a high component relying on technology. It included using computers, tablets, and phones by the students and teachers thought out all the lessons. Different activities were designed to be done online, such as the Prestest, the Posttest, quick games, and completing the Student Workbook. *Lesson 4.1 – Evaluating Context* was designed to be a self-study, and several other research artifacts, such as the bias confirmation questionnaire, the various exit tickets, the Student Survey, and the Demographic Questionnaire, were also designed to be online. Technology was crucial in delivering the CORE program.

Use of technology in the classroom. The analysis of the open-ended responses in all the data collection instruments revealed that using computers, tablets and phones during the intervention positively and negatively impacted class engagement and motivation. On the one hand, students appreciated using such technology for activities and assignments. They liked the convenience and efficiency of working digitally. On the other hand, researchers observed frequent students' distractions with students engaged in non-intervention activities, such as video games, movies, music, and browsing social media. Moreover, the researcher mentioned that some students misused computers, distracting not only the student themselves but also their peers, who were trying to see what they were doing, watching, or playing.

Technology Issues. The CORE program was designed with a high component relying on technology. Thus, presentation slides, activities to be done on the computer, a self-study module (i.e., *Lesson 4.1 – Evaluating Context*), the Pretest and Posttest, the Student Workbook, Student Survey, and the Demographic Questionnaire. Teachers experienced several events of technology failure, such as lack of WIFI, broken lesson content links, links not working properly, difficulty using the technology (such as the Jamboard), and audio issues for the self-study module. When technology failed, it disrupted the lesson flow, causing frustration among

teachers and students, student disengagement, and in some cases, like Chris' class when he did not have access to the Internet, loss of learning opportunities for the students. Nonetheless, it was also an opportunity for teachers to embrace adaptation and flexibility. They improvised, troubleshooted and solved, modified the instructional strategies, or put in place workarounds.

Data Collection Instruments

Research observers captured relevant information concerning three data collection instruments: Consent and Assent Forms, Demographic Questionnaire and the Classroom Observation Protocol.

Consent Assent Forms. Consent and assent forms are essential in interventions involving human participants, particularly for research in educational settings. They ensure participants are informed about the intervention's purpose, procedures, potential risks and benefits, and the participant's rights. The researchers noted that all teachers did not deliver these two forms during the first lesson. Moreover, the lesson plan did not incorporate or reserve time for the teacher to go over this form, thus impacting the effective time of the lesson class.

Demographic Questionnaire. A demographic questionnaire was used to collect information about the characteristics and backgrounds of students. It included questions about demographic factors such as age, gender, ethnicity, postal code, and other relevant variables. The questionnaire was not planned into lesson schedules; thus, the teacher asked their students to complete it at different times during the intervention. Some questions arose from the student regarding the collected data (e.g., gender, gender identification, and postal codes), impacting the lesson timing on which the teacher decided to collect this data.

Observation Protocol. The observation protocol designed to collect data through direct observation in the classroom during the intervention delivery was based on the BERI protocol. This protocol was designed to measure the motivation and engagement of large university classes, and it was adapted to be used in a high school classroom context. Some limitations were posed by the BERI Protocol when used by the researcher to collect engagement and

motivation behaviours. First, only ten students were observed at a time. Second, the researcher frequently recorded a lack of visibility on the students' behaviours and computer screens due to the observers' location. It was calculated from the Observation Protocol that 14% of the time, the observers recorded an "Unknown Engagement Score." Third, the observers lost control of the observed selected ten students since they moved around to perform the group activities.

Research Question 4: Program's Strengths and Weaknesses

The CORE intervention was developed to improve the ability of high school students in Quebec to evaluate online information critically. Like any approach or program, interventions have their strengths and weaknesses. Understanding these aspects can help better design, implement, and evaluate the intervention's effectiveness. Let us delve into the strengths and weaknesses of the CORE intervention.

CORE Program Strengths

1. Critical Thinking Skills: In today's digital age, information is abundant and easily accessible to anyone. Teaching students to discern between information and mis/disinformation promotes critical thinking. The CORE program content is a strength of this intervention since students learn to question, analyze and evaluate the sources and content of the information they encounter. This relevancy was validated by many comments participants mentioned in the Student's Engagement and Motivation Survey.
2. High Adherence to Content: The intervention was primarily delivered as expected by the research team, with some significant deviations or modifications. Since adherence to content is considered a mediating variable of the intervention outcomes, we can infer that it might positively impact the students' results. When the intervention content is delivered consistently across different school settings, the program becomes more reproducible, allowing future implementation to obtain similar outcomes.
3. Teachers' quality of delivery: Teachers obtained a very high score for the quality of the CORE program delivery. This score reflected not only their perspectives but the

research observers'. Within Carrol et al. (2007) implementation Fidelity Framework, quality of delivery is a potential moderator of the relationship between the intervention and the implementation fidelity. Thus, the high score for implementation fidelity might have positively been affected by the quality of delivery.

4. Evidence-based: The CORE intervention program was conducted based on research and best practices, providing a solid foundation for the effectiveness and success of the intervention. Implementing interventions with a proven track record increases the likelihood of achieving desired outcomes, if not during the first delivery, in subsequent implementation cycles.
5. Mixed method approach: Ample quantitative and qualitative evidence was rigorously gathered and analyzed from multiple sources or perspectives (teacher, student and researcher). Integrating quantitative and qualitative data gave us a more holistic and in-depth understanding of the CORE program. Quantitative data provided numerical insights and patterns, while qualitative data offered rich descriptions and contextual information. We triangulated data to validate the findings, enhancing the credibility and robustness of the conclusions.
6. CORE content with many group activities: Group activities were highlighted by most of the students as the most engaging features of the program. Participants were more motivated to actively participate when they were part of a dynamic group setting.

CORE Program Weaknesses

1. Missing Component Analysis: Breaking down an intervention or program into its essential components or elements is critical to the success of the intervention. The component analysis aims to provide a clear and detailed understanding of the intervention's most important aspects for effective implementation and impact. The lack of identifying essential components led teachers to adapt the content (skip or modify lessons or activities) that required special attention during the intervention execution and

should not have been altered.

2. Program Length: An evident and recurring theme observed from both teachers' and students' viewpoints was the extended duration of the intervention. According to their perspective, the intervention was excessively long, leading to students experiencing fatigue, boredom, and diminished interest. Simultaneously, teachers faced time constraints due to their other academic obligations, which further contributed to their need to make adjustments, including skipping certain content. Additionally, managing less engaged and motivated students required extra teacher effort, prompting them to adapt the intervention to address these challenges.
3. Overexposure to COVID-19: Students have been continuously exposed to discussions, news, and information about COVID-19 for an extended period. Exposing students to this topic during the intervention led to a sense of fatigue or exhaustion related to hearing about it repeatedly.
4. Observation Protocol: The data collection during the intervention delivery in the classroom involved using an observation protocol based on the BERI protocol. It was initially designed for measuring motivation and engagement in large university classes, but it was adapted for high school classroom use. However, the researcher encountered limitations while using the BERI Protocol to collect engagement and motivation behaviours. Firstly, only ten students could be observed at a time due to its design. Secondly, there were frequent instances where the observers had difficulty seeing the students' behaviours and computer screens because of their positioning. As a result, approximately 14% of the time, the observers recorded an "Unknown Engagement Score." Thirdly, as the selected ten students engaged in group activities, the observers lost control over monitoring their behaviours as they moved around.
5. Lack of Buffer Time: Buffer time refers to the extra time intentionally built into the program schedule to accommodate unexpected delays, challenges, or adjustments that

may arise during implementation. Teachers experienced various contextual factors and events that affected the intervention execution, such as student absenteeism, school activities, the timing of the lesson (a.m. or p.m.), and other students' assignments.

6. Delay in the Research Data Collection: Important data collection, such as consent and assent forms and demographic information, was not consistently gathered during the intervention since it was not accounted into the lesson plans. The delays resulted in negative consequences of research data loss.

Chapter Five: Discussion and Implications

This mixed methods study examined the fidelity of implementing a Critical Online Resource Evaluation (CORE) Program by evaluating the extent to which the intervention was delivered as planned regarding adherence to content, quality of delivery and participants' responsiveness. It used multiple sources (i.e., teachers, researchers and students) and methods to complement and corroborate data, providing a more comprehensive and robust understanding of the research findings. The purpose of the study was to provide the research team leaders, who are the primary stakeholders, with valuable insights on enhancing the quality and effectiveness of the CORE program so that they can make informed decisions regarding future intervention implementations. This chapter presents the results for each research question and discusses the emerging themes. According to the Utilization-Focus Evaluation Framework (UFE) used throughout this study, findings should be helpful to the program's primary stakeholders. Therefore, part of the discussion chapter is presented in a 'recommendation' format. Nonetheless, the results are also valuable for the field, so this chapter also covers considerations for future research.

The CORE program's fidelity level was high from a researcher's perspective, with a mean adherence to the content score of 89.0%. Thus, teachers closely adhered to the program as planned by the CORE team, delivering the entire intervention (nine lessons) on the estimated timeline and according to the school's regular classes schedule (75 minutes for ESL and 55 minutes for ELA.) From the teacher's perspective the level of adherence to content score was moderate (79.9%.) Attaining a high adherence and implementation fidelity level is challenging (Carroll et al., 2007). In the best scenario, programs are typically implemented with approximately 60% adherence to the intended protocol (Owczarzak et al., 2016; Dusenbury et al., 2003; Durlak & DuPre, 2008.) A review conducted by Durlak & DuPre (2008) concluded that few studies had achieved levels greater than 80%, and no study has reported a 100% implementation achievement. The CORE's program fidelity results are encouraging for the

primary stakeholders in their objective of continuing to grow the body of knowledge regarding the critical evaluation of online resources by implementing these interventions.

According to the implementation fidelity framework guiding this study, the implementation level achieved might be influenced by other variables: quality of delivery, participants' responsiveness, intervention complexity, and facilitation strategies (Carroll et al., 2007). This study found that the teachers delivered the CORE program with a high level of quality, with a mean score higher than 90% from both the researcher and the teacher's, suggesting they were well-prepared and confident to teach the content. Additionally, students were moderately engaged and motivated during the intervention, with a participants' responsiveness mean score of 77.01% from the researcher's perspective and 72.66% from the student's perspective. From the teachers' perspective, students' engagement and motivation were high, with a mean score of 80.9%. Intervention Complexity and Facilitation Strategies are two components of the implementation fidelity model adopted that were not part of the initial study scope. However, Intervention Complexity emerged during the qualitative analysis, along with four more emerging themes: Program Adaptation, Classroom-Level Independent Variables, Technology in the Classroom Setting, and Data Collection Instruments.

As a starting point for this discussion, we will refer to the CORE's program theory or theory of change (as called by some other authors), which made explicit the program's mechanisms to achieve its intended outcomes (Mowbray et al., 2003). The CORE intervention program has developed its theory of change by considering the impact of several factors on its success. These factors include the quality of the intervention material, the fidelity of the intervention, the student's engagement and motivation, and the teachers' preparation and confidence. The program hypothesizes that all of these elements play a role in determining the program's effectiveness. A logic model has been drawn to explicitly explain and graphically illustrate (Appendix A) how the program intended to achieve the proposed outcomes in the short, intermediate and long terms. For the first part of the program theory (i.e., *IF the research*

> *"IF the research team designed and developed a high-quality CORE intervention material AND the teachers delivered the lessons with fidelity, THEN high school students will improve their ability to evaluate online information critically, and more importantly, will be empowered to make better-informed decisions in their lives" (*Jimenez et al., 2021, p. 4*)*

team designed and developed a high-quality CORE intervention material AND the teachers delivered the lessons with fidelity), the study results indicate that teachers conducted the lessons with high fidelity, adhering to the intervention content, and delivering with a high level of quality. However, diverse opportunities were revealed to improve the material's quality and the intervention structure. The following sections will address these enhancement opportunities, first with general recommendations, followed by recommendations grouped by findings of each research question. The second part of the theory, thus "*high school students will improve their ability to evaluate online information critically,*" was addressed in a separate study conducted by the principal investigator of the CORE program.

General Recommendations

Conduct Component Analysis.

In the Conceptual Framework for Implementation fidelity (Figure 3), Carrol et al. (2007) outlined the five elements that must be measured for implementation fidelity and their relationship. One element that is part of the framework but was out of the scope of this study is *Program Differentiation*, to identify those essential components that are unique features, without which the CORE program will not have the intended effect. A clear example of essential components is the Pretest and Posttest. These two activities are key and must not be skipped. Moreover, teachers and observers shall pay special attention when students take these tests to ensure they are carried out smoothly (e.g., previous exhaustive testing done to ensure they work properly, the researcher's technical resources on standby during the test period). Another example observed during the intervention was that many teachers skipped the Confirmation

Bias Reflection Activity, which was stated as an important component by the primary stakeholders. During the component analysis, this element would be deemed as ‘essential’; therefore, teachers must not skip the activity.

Reduce the Intervention Length.

Students started to feel ‘fatigued’ by *Lesson 4.2 – Evaluating Source*, as mentioned by some of the teachers and corroborated by students. The starting point to reduce the intervention length is the component analysis activity. Once those elements are identified, other lesson parts can be minimized, removed or converted to extended activities. Students encountered repetition during the intervention; by identifying and eliminating that repetitive information or activities, students may keep their interest in the program.

Minimize Dependencies Between Lessons.

When possible, redesign the lessons with the least of dependencies. Teachers and students faced challenges derived from absences when the current lesson was based on the previous lesson’s activities (e.g., reading articles for the next class). In some cases, it might be necessary, such as the preparation for the Town hall (Lesson 2) and the Town hall (Lesson 3). In these cases, advise the teacher to plan for potential absences of team members (e.g., redistribution of teams and roles) and account for such events in the lesson timing.

Add Buffer Time to the Lessons.

Adding extra time to the lesson plan to account for unexpected events or unforeseen circumstances (e.g., car alarm off, other teacher interruption, technical issue, snow day) might benefit the teachers. By incorporating buffer time into a lesson plan, teachers can adapt the pace of the lesson without feeling rushed or compromising the timing or quality of the intervention. In addition, it contributes to smooth transitions between different activities or topics within a lesson, allows teachers to reinforce and review key concepts, and provide individualized support to students who may need it. In brief, buffer time allows teachers to adapt to their classroom context.

Align the Presentation Slides and the Student Workbook.

A thorough review must be conducted to make sure the presentation slides are aligned with the Student Workbook. It is important to maintain the alignment between these two artifacts to keep the flow of the lesson, promote student engagement, and keep a consistent and coherent learning experience.

Adapt the Instructional Content According to the Grade Level.

For future implementations of the CORE program, the content and structure must adapt to the grade level. Each grade level has specific learning objectives and standards that students are expected to achieve. The program should be mindful of this context, so teachers can effectively meet the unique needs of their students and promote engagement and motivation while achieving the CORE's objectives.

Recommendations Aiming at Improving Adherence to Content

Adherence in this study refers to whether the intervention was delivered as planned and adhered to the prescribed content and structure. It is the core measurement for program fidelity. The research community has had an ongoing debate between fidelity of implementation and adaptation within public health since the 1970s (Owczarzak et al., 2016; Carvalho et al., 2013; Durlak & DuPre, 2008). Such discussion revolves around the tension between fidelity to the original intervention model and the need to adapt the intervention to fit specific contexts or individual needs. The below recommendations by lesson are based on findings about how teachers adhere to the material, activities and flow of the lesson, what they changed, what they highlighted as engaging and relevant to the students, and what they skipped to accommodate to their context. These recommendations are framed under the 'component analysis' concept, and the purpose is to guide the CORE program re-design for subsequent implementations.

Lesson 1.1 – Intro & Pretest

- Keep the pizza indicators activity since it was one of the activities with a higher engagement score and has been described by the observers as a high engagement

activity.

- Label the Pretest as an ‘essential’ component.
- Account time for the Consent and Assent forms distribution and explanation.

Lesson 1.2 – Overview

- Keep the Belief Self-Assessment, Incredible Images and Looking for Indicators activities since the quantitative and qualitative findings suggest these are highly engaging activities.
- Label the Belief Self-Assessment activity as an ‘essential’ component.
- Combine *Lesson 1.1– Intro & Pretest* and *Lesson 1.2 – Overview* in one lesson to avoid repetition, as per recommended by some teachers.

Lesson 2 – Identifying a Problem

- Label the Town Hall Preparation activity as an essential component.
- Keep preparing the Town Hall activity since students were highly engaged with this activity. However, have an alternative plan for possible absenteeism of students for the next *Lesson 3 – Locating Information*, when the Town Hall activity takes place. Suppose only one student prepares and takes notes. In that case, the next lesson will encounter issues the teacher will need to deal with.
- Move the Town Hall Preparation activity to the beginning of the lesson before teaching the concepts. Kasey tried this approach and found it better since it allowed her to expand on how they found the sources. Amy (55) corroborated this idea; she followed the class flow, but then the explanation of the terminology took longer than expected, limiting the time for the Town hall preparation.
- Review the timing of the lesson. Some teachers run out of time during this lesson.
- Update the presentation slides to add the Confirmation of Bias concept in more detail, as Chris noted.

Lesson 3 – Locating Information

- Redesign the theoretical part of the lesson. In general, the lesson was more difficult to conduct as most of the teachers agreed. The theory was extensive and complicated.
- Add some terminology as an 'extended activity,' such as filter bubbles, echo chambers, query syntax, and algorithms.
- Keep Fake News and lateral reading concepts since those topics generated favourable student comments.
- Keep the Town Hall activity since it was one of the more engaging activities. However, move the activity to the first part of the lesson.
- Add the concepts of the definitions to the slides. This recommendation came from Amy (51) since she believed it easier to teach the unit with the definition on-screen (versus just being in the teacher's notes). Second, it was easier for the students to retain the information if they could read the definitions (versus just listening to her explain the terms without the definition on-screen).
- Allot more time for the Town Hall presentation. In Amy's (51) case, only five of the eight teams had time to present their arguments, so she had to take time during the next *Lesson 4.2 – Evaluating Context.*

Lesson 4.1 – Evaluating Context

- Consider redesigning this lesson as a non-self-study approach. The student did not show enthusiasm for activities that did not involve interaction with peers.
- If a decision is made to keep the lesson as is, resolve the lesson's technical difficulties (i.e., problems with loading Google Slides, navigating from a video to the next slide). Conduct a thorough testing of the lesson before the next round of interventions of the CORE program.

Lesson 4.2 – Evaluating Source

- Rethink the module considering engagement and motivation since this was the module

with the lowest student engaging score. In the redesign, consider a potential topic change (i.e., Covid) that could yield more positive results (See Figure 30 for students' topics of interest.)

- Replace Jamboard technology with a different whiteboard tool. Jamboard did not work well (e.g., the boards were shared among all teachers, and students could modify and even 'destroy' them). Moreover, some teachers mentioned not being familiar with it and having challenges to use it.
- Consider combining *Lesson 4.1 – Evaluating Context* and *Lesson 4.2. Evaluating Sou*rce, as suggested by Malena.

Lesson 4.3 – Evaluating Content

- Revaluate or eliminate some forms to fill out during this lesson: video, content evaluation, exit ticket.
- Rethink the activity of Preparation for Lesson 5. This activity is a homework activity that many students did not complete, creating challenges for the teacher during the following lesson 5 – *Synthesizing Information.* On a side note, this slide assumes the teacher will show it in presentation mode. If this is not the case (as it indeed happened), the images overlap, and the final effect is a distorted and confusing image.

Lesson 5 – Synthesizing Information & Lesson 6 – Communicating Findings

- Consider blending *Lesson 5 – Synthesizing Information* and *Lesson 6 – Communicating Findings*. The teacher, especially from grade 11th, skipped many activities in this lesson because their students had already learned these topics (i.e., prior knowledge). Several teachers proposed this modification, given that students had prior knowledge of the topic. Keep in mind, though, the timing of the class, since Malena performed this adaptation and resulted in negative consequences for the intervention results, such as lack of time to deliver the full content of the program appropriately, and more importantly, reducing the time students dedicated to the Posttest.

Independent Project

Make the independent project an 'essential' intervention component. If possible, this project should be graded. The independent project was optional based on the teacher's discretion and time available; therefore, the teachers skipped many activities related to the independent project. Incorporate activities related to the Independent Project since the beginning of the intervention. For example, define the topic in the first lesson (as homework for the next class), locate sources after lesson 2, and so on. As a fair adaptation of the program to the class context, teachers might be allowed to adjust the project to their regular class, as was the case of the 'persuasive paper' for Kasey and Malena.

Recommendation Aiming at Improving Quality of Delivery

The quality of the delivery of an intervention can serve as a moderator in the relationship between the intervention and its implementation fidelity. This concept refers to how the implementer carries out the program, including recommended techniques, processes, and methods. If the intervention is poorly delivered, it can negatively impact the implementation fidelity level. The quality of delivery score was very high (93.5%); therefore, only two recommendations arose.

Conducting the intervention for the second time.

While recruiting teachers posed challenges for the CORE program, it is advisable to attempt to enlist the same teachers for the program's second implementation. For example, Amy's (51) group was the second group to receive the CORE intervention, and the adherence to content had a better score than Amy's (55). This difference can be explained by the fact that the teacher felt more confident and prepared the second time and had the opportunity to accommodate the intervention material to her needs based on previous experience. Also, note that the lack of time to complete the Town Hall activity could be due to other factors. For example, as teachers become more familiar with the material, they become more efficient at teaching the lessons and can finish on time. When teaching the lessons for the first few times,

they may have a challenge getting through all the material.

Train Research Observers.

Some of the researchers' notes regarding observations about teachers' behaviours related to preparation and confidence (quality of delivery were insufficient and lacking details. According to the Jotform blog, there are a few basic questions that could help observers capture behaviours that demonstrate the teacher's quality of delivery:

- Is the teacher able to answer student questions accurately?
- Is the teacher able to elaborate on topics and incorporate their knowledge outside of the intervention?
- Does the teacher make the subject matter relatable and incorporate real-world examples?
- Does the teacher speak clearly and loud enough for everyone in the room to hear?
- Is the teacher checking for understanding throughout the lesson?
- Does the teacher respond to questions with patience and understanding?
- Does the teacher leave room for questions before completing the lesson?
- Does the teacher keep distractions to a minimum by handling any disruptive students?
- Does the class feel comfortable enough for students to speak up and be heard?
- Does the teacher use inclusive and inoffensive language?

Recommendations Aiming at Improving Participant's Responsiveness

Student engagement and motivation was statistically significantly different between different classrooms, F (3, 91) = 3.320, p = .023, ω^2 = 0.068. Student engagement score increased from Kasey's (M = 67.74, SD = 11.70) to Amy's (M = 68.95, SD = 18.86), Malena's (M = 69.42, SD = 12.58) and Chris' (M = 84.55, SD = 10.46) classrooms, in that order. Student engagement and motivation is a complex construct determined by various individual and contextual factors. Attempting to comprehensively explain student engagement results is out of the scope of this evaluation. However, several plausible explanations for Chris' case, reporting

a significant difference in student engagement compared to other cases, can be hypothesized. One relevant factor might be grade-level differences. Chris' class is 8th grade, the lowest grade level of the five participant cases. Evidence suggests that student engagement decreases as they advance from upper elementary to middle school, with the lowest levels observed in high school (Fredricks et al., 2011; Marks, 2000). Another reason that could explain the significant differences is that the school is private in Chris's case. Furthermore, the CORE Pretest and Posttest results suggested that the intervention positively affected private school students (Corrigan et al., 2022). A third factor that might influence the results is the subject matter. According to Marks (2000), class subject matter impacts student engagement in high school students. Whether History class versus English as a Second Language or English Language Arts influenced the significant difference in Chris' case engagement scores is a matter of further research. Chris adapted several CORE activities and examples to his contextual History class. In summary, it is crucial to consider a holistic approach when examining student engagement and motivation concerning other significant factors and their interactions.

The following recommendations are intended to inform the instructional re-design of the CORE content for upcoming implementations related to student engagement and motivation based on teacher, student, and researcher findings.

Rethink the Activities Approach – Group Activities.

Whenever possible, add group activities or replace individual to group activities to allow students to interact with their peers, collaborate, and engage in discussions.

Consider Topics other than Covid.

To avoid student fatigue and boredom with the Covid topic, propose other topics that raise the student's interest., such as politics, social justice, science, discrimination, sports, education, climate change, and mental health, as mentioned in the Student's Engagement and Motivation Survey.

Implement the Exit Ticket at the End of all Lessons.

Replace the Student's Engagement and Motivation Survey by implementing 'exit tickets' at the end of each lesson. These short assessments or reflections that students complete before leaving the lesson provide more granular and valuable data for the researcher's purposes. Teachers will also benefit from this information since they can identify areas that need reinforcement and patterns in student understanding and gauge the engagement and motivation of the students.

Recommendations Aiming at Improving Emerging Themes

Intervention Complexity

Recommendations about intervention complexity have been highlighted in this chapter (see Recommendations Aiming to Improve Adherence to Content.) However, we want to make a special mention of redesigning *Lesson 3 - Locating Information.* Teachers found this lesson to be one of the most difficult and demanding. The concepts and terms were challenging to explain, requiring additional preparation before class and more in-depth explanations during teaching.

Technology in the Classroom Setting

Allows the use of Computers, Laptops and Cellphones. Although technology is indeed a source of distraction, not only for the student engaged in the non-class activity (i.e., video game, movie, social networking) but the peers, it is an integral part of the design of the CORE program. Students appreciated such features. Teachers can implement several effective classroom management strategies to allow the use of technology and minimize distractions, such as setting clear expectations and boundaries with the students, establishing a mechanism to supervise and monitor student's screens, randomly selecting students to participate in the discussion, have students to put away their laptops during presentation time and only open them when they have to do productive work.

Mitigate the Impact of Technology Failures. To alleviate the effects of technology

issues in the classroom. Teachers need to have backup plans in place. This preparation could include having alternative activities or resources that can be used offline, incorporating non-digital teaching methods (e.g. printed material), or having contingency plans for accessing technology in different locations or using different devices.

Data Collection Instruments

Account Time to Complete Consent and Assent Forms during First Lesson. Teachers dedicated time explaining to the students the participation assent and consent forms. This time should be accounted for in the Lesson timing plan. Once the forms are delivered to the students, the research team must conduct a thorough control and follow-up to ensure the maximum number of students enrol in the intervention. Without proper consent, the integrity and validity of the research or intervention may be compromised.

Account for Time to Complete the Demographic Questionnaire During the First Lesson. Different teachers delivered the demographic questionnaire during different lessons, affecting the timing of the lesson and the intervention. Students asked questions regarding some information (e.g., what CORE stands for, gender options). Teachers dedicated a good amount of time to explanations. This Demographic Questionnaire and the Pretest should be completed during the first lesson, and time must be accounted for on the lesson plan.

Rethink / Redesign Observation Protocol. There were several reasons why observers could not measure the student engagement and motivation, such as group activities forced the students to move around, and frequently students did not come back to their places; depending on the observer's position, sometimes it was not possible to know what the students were doing in their computers, some activities were too short, and there was no time to collect the data. According to the recommendation to increase the group activities throughout the entire intervention, the (BERI-based) Observation Protocol might not be the best suited for capturing observational data during the intervention.

Chapter Six: Limitations and Conclusion

Below, we describe some of the limitations encountered doing this study regarding the research design, sampling and instruments used to collect data.

First, despite the positive aspects of using mixed methods research (MMR), there are some critical drawbacks, mainly when it is being conducted to fulfil a postgraduate degree (Halcomb, 2018; Regnault et al., 2018). The major limitation of the MMR was the high demand for methodological skillsets and time. It required a broad range of research skills and experience beyond quantitative and qualitative. Integrating qualitative and quantitative data in the analysis required additional time, maybe longer than initially planned.

Second, as basic research, program evaluation focuses on generalizing findings (Walser & Trevisan, 2015). In contrast, this program evaluation study was conducted as applied research focused on supporting decision-making about the CORE program in that context at that time. Moreover, when applying the same intervention to different school settings, every class's context and experiences differ (Hagerman, 2019). In this sense, the nature of this program evaluation poses a limitation in the generalization of the findings.

Third, the CORE program used convenience sampling to recruit four teachers who responded to the call for participants. This was a small sample size for quantitative analysis on adherence and quality of delivery, limiting the generalization of the results. However, using mixed methods and data triangulation allowed us to deepen the analysis and understanding of the extent to which the program was implemented with fidelity, the experiences and the challenges.

Four, the evaluator and the research team decided on the questions, the instruments to be used, and the data analysis methods. When only evaluators and clients get to decide on the evaluation design, "other stakeholders will be denied the opportunity to pursue their legitimate interest" (Guba & Lincoln, 1989, p. 9). This selective involvement might not account for other stakeholders' (e.g., teachers') legitimate interests.

Lastly, regarding the data collection, there are few limitations to consider. In regards to the BERI protocol, it has been demonstrated to be reliable for collecting classroom observation engagement behaviours. “Data from 2,154 judgments of individual student engagement, from six pairs of observers in three different educational settings, with five different instructors were used to evaluate interrater reliability. The average interrater agreement was calculated to be 96.5%” (Lane & Harris, 2015, p 87). However, one concern regarding the use of observation protocols is that the students may change their behaviour when being observed, particularly in these school settings where the participants were aware of the observer’s presence and note recording; this might imply the gathered evidence and what type of claims can be made (CDC, 1999). With respect to the Student’s Engagement and Motivation Survey, even though it complies with the validity and reliability standards with a reported Cronbach's alpha of .65 to .68, it is a self-reported instrument. Self-reports rely on the students' responses about their motivational beliefs and attitudes (Fredricks & McColskey, 2012) and can be subject to biases, such as social desirability; students are more likely to report experiences that are considered socially acceptable or preferred (Pekrun, 2020). Likewise, the Teacher's Journal Log relies on self-reporting and might pose similar limitations to the student motivation and engagement survey.

Concluding Remarks and Future Directions for Research

Information gathered from implementing interventions in schools is extremely valuable for evaluating their effectiveness. It can reveal insights into enhancing current and future implementation processes by measuring outcomes like fidelity. This study leads to a handful of significant conclusions that can be reasonably inferred. First, the fidelity scores across the CORE intervention project were higher than expected. These findings aligned with existing literature, which indicates that achieving fidelity scores near 100% is unrealistic (Carroll et al., 2007), even when implementation is carefully planned and executed. Second, this research produced evidence that teachers made adjustments to the intervention content while

implementing it, taking into account specific local contexts such as grade level, language of instruction, class subject matter (e.g., ESL, ELA, History), previous content knowledge, or unanticipated changes in available resources, like computer access. These modifications seemed to be the result of teachers' meaningful professional judgement. These findings align with certain aspects of the existing literature (e.g., Fullan, 2001) that educators have a crucial professional duty to be responsive to the dynamic and evolving demands of the classroom and real-world situations. Third, this evaluation produced results and insights for the primary stakeholders that will be useful and practical to incorporate in the next round of the CORE intervention implementation. Recommendations were provided for redesigning the content material, the structure of the intervention, the topics and the instructional design with the intent to improve the program fidelity, quality of delivery, and student engagement and motivation, with an ultimate goal of improving students' ability to evaluate online information critically.

This study was limited in its assessment of student outcomes. Examining intervention outcomes alongside implementation outcomes helps us understand why the effectiveness of interventions varies across different schools and teacher settings (Schultes, 2021.) "Programs that are implemented with high levels of fidelity but fail to produce desired effects may need to be redesigned" (Dusenbury et al., 2003, p. 240). Analyzing CORE intervention outcomes in combination with CORE implementation outcomes would enable a potentially more insightful and accurate interpretation of findings and the reasons for variabilities in the intervention's effectiveness in different schools and teachers. This research has important implications for future studies and the broader field of program fidelity in terms of providing necessary information for assessing fidelity methods in the context of classroom interventions. It contributes valuable data to the currently limited evidence base regarding the application of fidelity assessment methods in this area. Additionally, we have demonstrated the effectiveness of employing a combination of quantitative and qualitative data through a mixed methods approach, leading to a deeper comprehension of the variables that impact fidelity.

References

Alkin, M. C., & Vo, A. T. (2017). *Evaluation essentials, second edition: From A to Z*. The Guilford Press.

An, M., Dusing, S. C., Harbourne, R. T., & Sheridan, S. M. (2020). What Really Works in Intervention? Using Fidelity Measures to Support Optimal Outcomes. *Physical Therapy*, *100*(5), 757–765. https://doi.org/10.1093/ptj/pzaa006

Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation*, *33*, 5–17.

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: analysis of a complex cognitive skill. *Computers in Human Behavior*, *21*(3), 487–508. https://doi.org/10.1016/j.chb.2004.10.005

Breitenstein, S. M., Gross, D., Garvey, C. A., Hill, C., Fogg, L., & Resnick, B. (2010). Implementation fidelity in community-based interventions. *Research in Nursing &Amp; Health*, *33*(2), 164–173. https://doi.org/10.1002/nur.20373

Bryson, J. M., & Quinn Patton, M. (2015). Analyzing and engaging stakeholders. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th Ed.). Hoboken, NJ: Jossey-Bass & Pfeiffer Imprints, Wiley. Retrieved from https://concordiauniversity.on.worldcat.org/oclc/905668071

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*(1). https://doi.org/10.1186/1748-5908-2-40

Carvalho, M., Honeycutt, S., Escoffery, C., Glanz, K., Sabbs, D., & Kegler, M. C. (2013). Balancing Fidelity and Adaptation. *Journal of Public Health Management and Practice*, *19*(4), 348–356. https://doi.org/10.1097/phh.0b013e31826d80eb

CDC. (1999). *Framework for program evaluation in public health* (RR-11). Centers for Disease

Control and Prevention, Program Performance and Evaluation Office. https://www.cdc.gov/mmwr/PDF/rr/rr4811.pdf

Ceci, L. (2023, January 9). *Most downloaded apps worldwide 2022 | Statista*. Statista. https://www.statista.com/statistics/1285960/top-downloaded-mobile-apps-worldwide/

Clark, P. V. L., & Creswell, J. W. (2014). *Understanding Research: A Consumer's Guide* (2nd ed.). Pearson.

Coiro, J., Coscarelli, C., Maykel, C., & Forzani, E. (2015). Investigating criteria that seventh graders use to evaluate the quality of online information. *Journal of Adolescent & Adult Literacy*, *59*(3), 287–297. https://doi.org/10.1002/jaal.448

Combs, K. M., Buckley, P. R., Lain, M. A., Drewelow, K. M., Urano, G., & Kerns, S. E. U. (2022). Influence of Classroom-Level Factors on Implementation Fidelity During Scale-up of Evidence-Based Interventions. *Prevention Science*, *23*(6), 969–981. https://doi.org/10.1007/s11121-022-01375-3

Corrigan, J. (2019). *Writing in times of deixis: A validation study of a large-scale assessment of New Literacies* [Doctoral dissertation]. University of Ottawa. https://ruor.uottawa.ca/bitstream/10393/38696/15/Corrigan_Julie_A_2018_thesis.pdf

Corrigan, J., McDonough, K., Forzani, E., Neumann, H., & Jimenez, M. (2022). *"The Publisher is Biased": Evaluating a Teacher-Led Intervention to Improve Students' Online Evaluation Practices*. [Paper presentation]. 2023 AERA Annual Meeting, Chicago, IL, , United States of America.

Dixon, S. (2022, July 15). *Social media usage in Canada - statistics & facts*. Statista. https://www.statista.com/topics/2729/social-networking-in-canada/#dossierKeyfigures

Durlak, J. A., & DuPre, E. P. (2008). Implementation Matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research*, *18*(2), 237–256. https://doi.org/10.1093/her/18.2.237

Fetters, M. D. (2020). The Mixed Methods Research Workbook: Activities for Designing, Implementing, and Publishing Projects. https://doi.org/10.4135/9781071909713

Fleming, S. (2019, September 2). *This is how many websites exist globally*. World Economic Forum. https://www.weforum.org/agenda/2019/09/chart-of-the-day-how-many-websites-are-there/

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2015). *ICILS 2013 technical report*. International Association for the Evaluation of Educational Achievement. https://www.iea.nl/studies/iea/icils/2013

Fredricks, J. A., & McColskey, W. (2012). The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments. *Handbook of Research on Student Engagement*, 763–782.

Fredricks, J., McColskey, W., Meli, J., Montrosse, B., Mordica, J., & Mooney, K. (2011). *Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments* (Issues&Answers Report, REL 2011–No. 098). Regional Educational Laboratory Program. https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_2011098.pdf

Fullan, M. (2001). The new meaning of educational change. In *Routledge eBooks*. https://doi.org/10.4324/9780203986561

Golden, G., & Toledo Figueroa, D. (2020). *Education policy evaluation – Surveying the OECD landscape*. Organisation for Economic Co-operation and Development. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/WKP(2020)24&docLanguage=En

Guba, E., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation* (First). Sage Publications.

Hagerman, M. S. (2019). Digital Literacies Learning in Contexts of Development: A Critical Review of Six IDRC-Funded Interventions 2016–2018. *Media and Communication*, *7*(2), 115–127. https://doi.org/10.17645/mac.v7i2.1959

Halcomb, E. J. (2018). Mixed methods research: The issues beyond combining methods. *Journal of Advanced Nursing*, *75*(3), 499–501. https://doi.org/10.1111/jan.13877

Hämäläinen, E. K., Kiili, C., Marttunen, M., Räikkönen, E., González-Ibáñez, R., & Leppänen, P. H. (2020). Promoting sixth graders' credibility evaluation of web pages: An intervention study. *Computers in Human Behavior*, *110*, 106372. https://doi.org/10.1016/j.chb.2020.106372

Hatry, H. P., Newcomer, K. E., & Wholey, J. S. (2015). *Handbook of Practical Program Evaluation* (4th ed.). Jossey-Bass.

Jimenez, M., Durocher, C., & Van-Patter, D. (2021). *Critical online resources evaluation (CORE) intervention program: Implementation evaluation plan* (Unpublished paper). Department of Education, Concordia University.

Kanniainen, L., Kiili, C., Tolvanen, A., Aro, M., & Leppänen, P. H. T. (2019). Literacy skills and online research and comprehension: struggling readers face difficulties online. *Reading and Writing*, *32*(9), 2201–2222. https://doi.org/10.1007/s11145-019-09944-9

Kiili, C., Laurinen, L., & Marttunen, M. (2008). Students evaluating internet sources: From Versatile Evaluators to Uncritical Readers. *Journal of Educational Computing Research*, *39*(1), 75–95. https://doi.org/10.2190/ec.39.1.e

Kiili, C., Leu, D. J., Marttunen, M., Hautala, J., & Leppänen, P. H. T. (2017). Exploring early adolescents' evaluation of academic and commercial online resources related to health. *Reading and Writing*, *31*(3), 533–557. https://doi.org/10.1007/s11145-017-9797-2

Lane, E., & Harris, S. (2015). Research and Teaching: A New Tool for Measuring Student Behavioral Engagement in Large University Classes. *Journal of College Science Teaching*, *044*(06). https://doi.org/10.2505/4/jcst15_044_06_83

Leu, D., Kinzer, C. K., Coiro, J., & Cammack, D. W. (2004). Toward a Theory of New Literacies

Emerging From the Internet and Other Information and Communication Technologies. *Theoretical Models and Processes of Reading*, 1570–1613. https://doi.org/10.1598/0872075028.54

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2017). New Literacies: A Dual-Level Theory of the Changing Nature of Literacy, Instruction, and Assessment. *Journal of Education*, *197*(2), 1–18. https://doi.org/10.1177/002205741719700202

Leu, D., Kinzer, C. K., Coiro, J., & Cammack, D. W. (2004). Toward a Theory of New Literacies Emerging From the Internet and Other Information and Communication Technologies. *Theoretical Models and Processes of Reading*, 1570–1613. https://doi.org/10.1598/0872075028.54

Lund, A., & Lund, M. (2020). *SPSS Statistics Tutorials and Statistical Guides*. Laerd Statistics. https://statistics.laerd.com/

Macedo-Rouet, M., Potocki, A., Scharrer, L., Ros, C., Stadtler, M., Salmerón, L., & Rouet, J. (2019). How good is this page? Benefits and limits of prompting on adolescents' evaluation of web information quality. *Reading Research Quarterly*, *54*(3), 299–321. https://doi.org/10.1002/rrq.241

Marks, H. M. (2000). Student Engagement in Instructional Activity: Patterns in the Elementary, Middle, and High School Years. *American Educational Research Journal*, *37*(1), 153–184. https://doi.org/10.3102/00028312037001153

Mathison, S. (Ed.) (2005). *Encyclopedia of evaluation*. (Vols. 1-0). Sage Publications, Inc., https://dx.doi.org/10.4135/9781412950558

Mathison, S. (2008). What is the difference between evaluation and research—and why do we care? In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 183–196). Guilford Press

McGrew, S. (2020). Learning to evaluate: An intervention in civic online reasoning. *Computers & Education*, *145*, 103711. https://doi.org/10.1016/j.compedu.2019.103711

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory &Amp; Research in Social Education*, *46*(2), 165–193. https://doi.org/10.1080/00933104.2017.1416320

McLaughlin, J. A., & Jordan, G. B. (2015). Using logic models. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed.). Jossey-Bass & Pfeiffer Imprints, Wiley.

Mertens, D. M. (2015). Philosophical assumptions and program evaluation. *Enrico Guglielminetti Luciana Regina*, *75*, 75-85.

Ministère de l'Éducation et de l'Enseignement Supérieur. (2018). *Digital Action Plan for Education and Higher Education*. Retrieved April 28, 2021, from http://www.education.gouv.qc.ca/en/references/tx-solrtyperecherchepublicationtx-solrpublicationnouveaute/results/detail/article/digital-action-plan-for-education-and-higher-education/?a=a&cHash=7a054beb11368e142830bc08c55347f4

Miles, M. B., & Huberman, M. A. (1994). *Qualitative Data Analysis: An Expanded Sourcebook, 2nd Edition* (2nd ed.). SAGE Publications, Inc.

Mittermeyer, D. (2005). Incoming first year undergraduate students: How information literate are they? *Education for Information*, *23*(4), 203–232. https://doi.org/10.3233/efi-2005-23401

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity Criteria: Development, Measurement, and Validation. *American Journal of Evaluation*, *24*(3), 315–340. https://doi.org/10.1177/109821400302400303

O'Donnell, C. L. (2008). Defining, Conceptualizing, and Measuring Fidelity of Implementation and Its Relationship to Outcomes in K–12 Curriculum Intervention Research. *Review of Educational Research*, *78*(1), 33–84. https://doi.org/10.3102/0034654307313793

Owczarzak, J., Broaddus, M. R., & Pinkerton, S. D. (2016). A qualitative analysis of the concepts of fidelity and adaptation in the implementation of an evidence-based HIV

prevention intervention. *Health Education Research*, *31*(2), 283–294. https://doi.org/10.1093/her/cyw012

Owen, J. M. (2006). *Program evaluation: Forms and approaches* (3rd ed.). The Guilford Press.

Patton, M. Q. (1994). The program evaluation standards: How to assess evaluations of educational programs. *Evaluation Practice*, *15*(2), 193–199. https://doi.org/10.1016/0886-1633(94)90009-4

Patton, M. Q. (2008). *Utilization-Focused Evaluation* (4th ed.). Sage Publications, Inc.

Pekrun, R. (2020). Self-Report is Indispensable to Assess Students' Learning. *Frontline Learning Research*, *8*(3), 185–193. https://doi.org/10.14786/flr.v8i3.637

Pérez, A., Potocki, A., Stadtler, M., Macedo-Rouet, M., Paul, J., Salmerón, L., & Rouet, J. F. (2018). Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions. *Learning and Instruction*, *58*, 53–64. https://doi.org/10.1016/j.learninstruc.2018.04.006

Ramirez, R., & Brodhead, D. (2013). *Utilization focused evaluation: A primer for evaluators.* https://www.researchgate.net/publication/263580952

Regnault, A., Willgoss, T., & Barbic, S. (2018). Towards the use of mixed methods inquiry as best practice in health outcomes research. *Journal of Patient-Reported Outcomes*, *2*(1). https://doi.org/10.1186/s41687-018-0043-8

Schultes, M. (2021). An introduction to implementation evaluation of school-based interventions. *European Journal of Developmental Psychology*, *20*(1), 189–201. https://doi.org/10.1080/17405629.2021.1976633

Sidani, S., & Sechrest, L. (1999). Putting Program Theory into Operation. *American Journal of Evaluation*, *20*(2), 227–238.

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2011). Intervention Fidelity in Special and General Education Research Journals. *The Journal of Special Education*, *47*(1), 3–13. https://doi.org/10.1177/0022466911419516

Toomey, E., Matthews, J., & Hurley, D. A. (2017). Using mixed methods to assess fidelity of delivery and its influencing factors in a complex self-management intervention for people with osteoarthritis and low back pain. *BMJ Open*, *7*(8), e015452. https://doi.org/10.1136/bmjopen-2016-015452

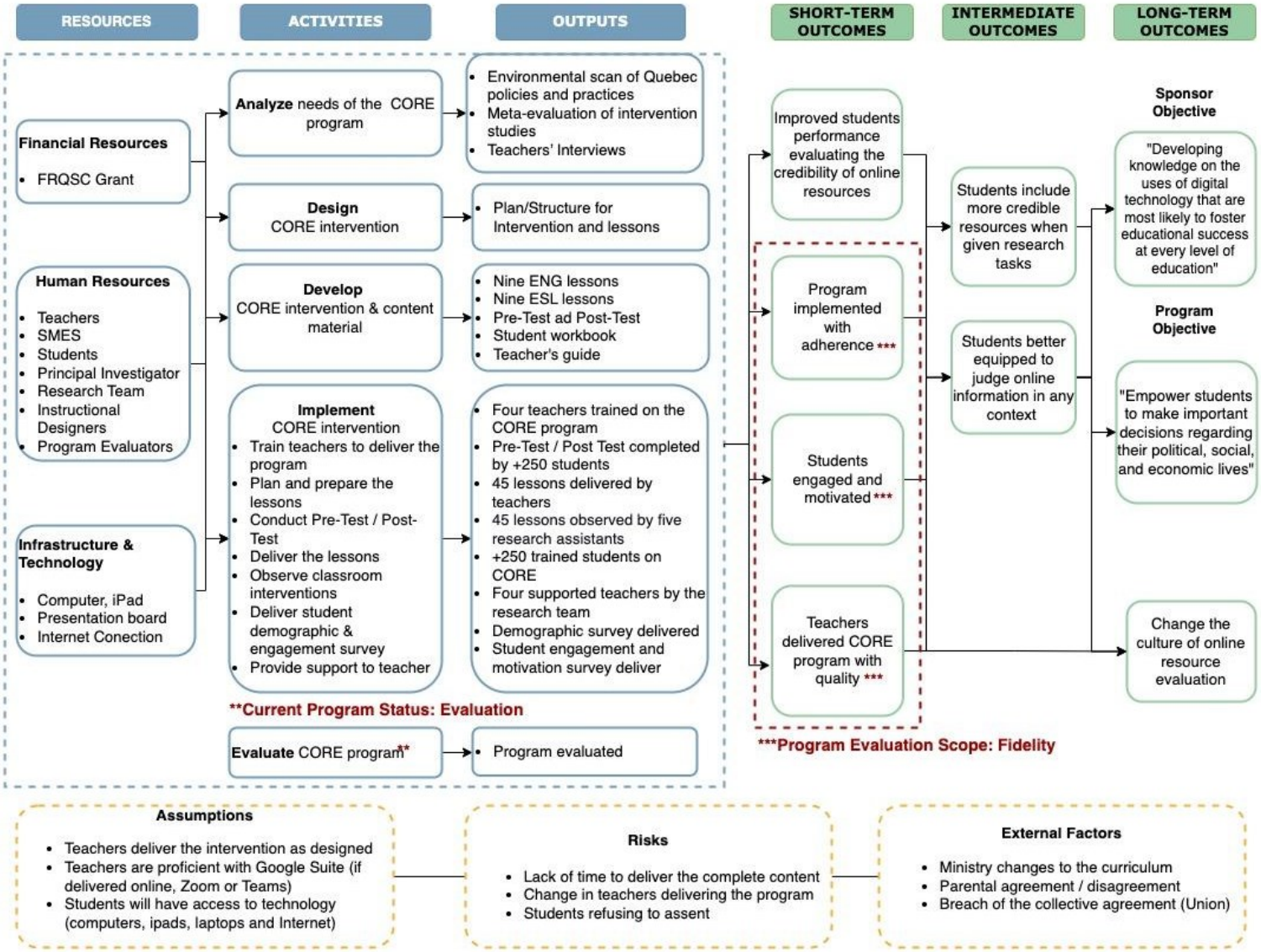
Vedung, E. (2000). *Public policy and program evaluation*. Transaction.

Walser, T. M., & Trevisan, M. S. (2015). Evaluability Assessment Thesis and Dissertation Studies in Graduate Professional Degree Programs. *American Journal of Evaluation*, *37*(1), 118–138. https://doi.org/10.1177/1098214015583693

Yarbrough, D.B., Shula, L.M., Hopson, R.K., & Caruthers, F.A. (2010). *The Program Evaluation Standards : A guide for evaluators and evaluation users* (3rd. ed). Thousand Oaks, CA: Corwin Press. https://jcsee.org/program/

Zhang, S., & Duke, N. K. (2011). The impact of instruction in the WWWDOT framework on students' disposition and ability to evaluate web sites as sources of information. *The Elementary School Journal*, *112*(1), 132–154. https://doi.org/10.1086/660687

Appendix A – CORE LOGIC MODEL (FULL VERSION)



Appendix B – Student Engagements Scores by Activity

Lesson	Part	Kasey	Malena	Amy (51)	Amy (55)	Chris	M
1.1 Intro & Pretest	Intro	92.86	100.00	100.00	92.86	87.50	94.64
	Warm-Up: Pizza Indicators	100.00	100.00	79.31	77.78	80.95	87.61
	Pretest	100.00	100.00	100.00	100.00	80.00	96.00
1.2 Overview	Intro	66.67	84.21	100.00	50.00	100.00	80.18
	Warm-Up: Incredible Images	100.00	76.00	76.47	100.00	Skipped	88.12
	Part1: Looking for Indicator Activity	100.00	91.67	86.21	50.00	100.00	85.58
	Part 2: Sharing Indicators Activity	100.00	92.86	33.33	N/A	86.67	78.22
	Part 3: Ranking Indicators Activity	100.00	100.00		N/A	73.68	91.23
	Part 4: Experts Knowledge: Three Tier Framework	100.00	92.86	46.15	95.00	23.08	71.42
	Part 5: Exit Ticket	Skipped	76.92	Skipped	N/A	Skipped	76.92
2. Identifying a Problem	Intro	100.00	0.00	93.10	100.00	77.78	74.18
	Part 1: Beliefs Self-Assessment	100.00	100.00	96.67	100.00	58.82	91.10
	Part 2: Terminology Lecture	100.00	100.00	86.67	90.00	50.00	85.33
	Part 3: Town Hall Preparation	100.00	92.86	86.67	100.00	47.83	85.47
	Part 3: Bias Confirmation Reflection	Skipped	100.00	Skipped	Skipped	61.54	80.77
3. Locating Information	Intro	48.15	N/A	80.00	83.33	31.25	60.68
	Part 1: Concepts Lecture	50.00	70.83	76.67	86.67	70.37	70.91
	Part 2: Concepts Lecture & Activity	43.33	64.29	76.67	83.33	64.00	66.32
	Part 3: Lateral Reading	46.67	55.00	76.67	73.33	70.83	64.50
	Part 4: Town Hall Activity	70.00	35.71	80.95	90.00	93.33	74.00

Lesson	Part	Kasey	Malena	Amy (51)	Amy (55)	Chris	M
4.1 Evaluating Context	Intro	100.00	94.44	62.50	70.00	71.43	79.67
	Context Self Study Module	100	96.30	N/A	37.50	72.22	76.51
	Closing - Assessment on Lesson	60.00	N/A	N/A	N/A	87.50	73.75
4.2 Evaluating Source	Intro	72.22	63.64	77.78	75.00	50.00	67.73
	Part 1: Evaluating Source Credibility Lecture	77.78	100.00	75.86	73.08	48.00	74.94
	Class Activity 1: Four Corners	53.33	100.00	93.33	77.78	50.00	74.89
	Types of Organizations Lecture	67.86	57.14	83.33	Skipped	57.14	66.37
	Class Activity 2: CBS Video	80.00	69.23	68.97	N/A	55.56	68.44
	Web of Understanding Lecture	50.00	N/A	62.50	Skipped	28.57	47.02
	Author's Expertise Lecture	57.89	100.00	70.00	N/A	41.67	67.39
	Part 2: Taking a Cririca l Stance Activity	73.33	100.00	76.67	87.50	78.57	83.21
	Independent Project: Narrowing Down Resources	Skipped	Skipped	Skipped	75.00	Skipped	75.00
4.3 Evaluating Content	Intro	100.00	100.00	54.55	N/A	83.33	84.47
	Warm-Up: Dr. Tenpenny Video	64.71	100.00	70.59	76.47	50.00	72.35
	Fact-checking with Snopes	100.00	100.00	20.00	N/A	50.00	67.50
	Learning Objectives	100.00	100.00	N/A	N/A	N/A	100.00
	Part 1:Evaluating Content Credibility Lecture	90.00	94.74	34.48	58.33	60.87	67.68
	Part 2: Taking a Critical Stance Activity	67.86	100.00	70.00	50.00	63.16	70.20
	Summary	Skipped	Skipped	Skipped	57.14	28.57	42.86
	Lesson 5 Preparation	100.00	100.00	100.00	N/A	42.86	85.72
	Part 3: Independent Project	Skipped	100.00	N/A	N/A	N/A	100.00
	Part 4: Exit Ticket	100.00	100.00	N/A	Skipped	66.67	88.89

Lesson	Part	Kasey	Malena	Amy (51)	Amy (55)	Chris	M
5. Synthesizing Information	Intro	N/A	86.36	94.44	55.00	29.63	66.36
	Part 1: Understanding the concepts	100.00	83.33	70.37	52.00	33.33	67.81
	Concept Map: General	100.00	93.75	N/A	N/A	47.06	80.27
	Concept Map: Example A	100.00	Skipped	100.00	N/A	N/A	100.00
	Part 2: Activity 1 Organize	90.91	75.86	62.50	50.00	59.26	67.71
	Part 2: Activity 2 Compare	Skipped	Skipped	N/A	Skipped	100.00	100.00
	Part 3: Connect/ Concept Mapping	100.00	Skipped	62.50	N/A	73.08	78.53
	Individual Concept Map	100.00	95.45	Skipped	Skipped	89.29	94.91
	Part 4: Independent Project: Evaluating Resources	Skipped	Skipped	62.50	75.00	N/A	68.75
6. Communicating Findings	Intro	78.57	91.67	60.00	70.00	30.77	66.20
	Concept map from lesson 5	77.78	Skipped	80.00	66.67	50.00	68.61
	Part 1: Writing a Thesis Statement	75.00	Skipped	80.00	68.18	19.23	60.60
	Posttest	71.43	100.00	N/A	N/A	100.00	90.48

Note. N/A: Not measured by the observer; Skipped: the teacher omitted the activity

Appendix C – ANOVA Supporting Documentation

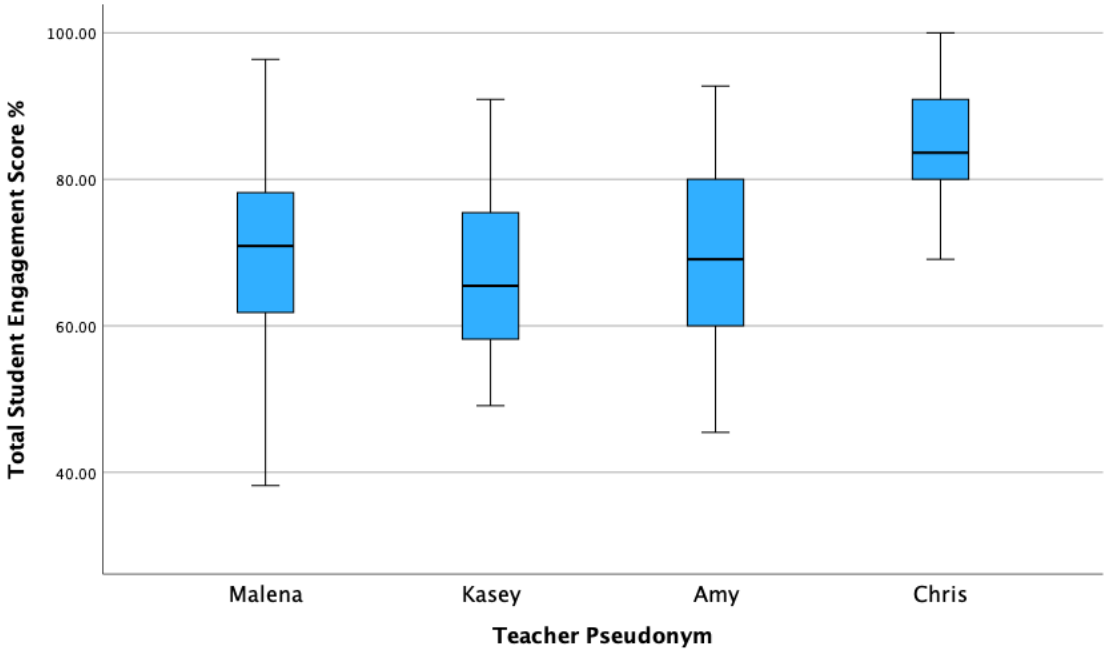
A one-way analysis of variance (ANOVA) was used to determine whether there were any statistically significant differences between the student and engagement and motivation means of the four independent groups (i.e., classrooms/interventions). The dependant continuous variable was the scores of student engagement and motivation and the independent variable was the classroom (i.e., Amy, Chris, Kasey and Malena).

Assumptions Validation

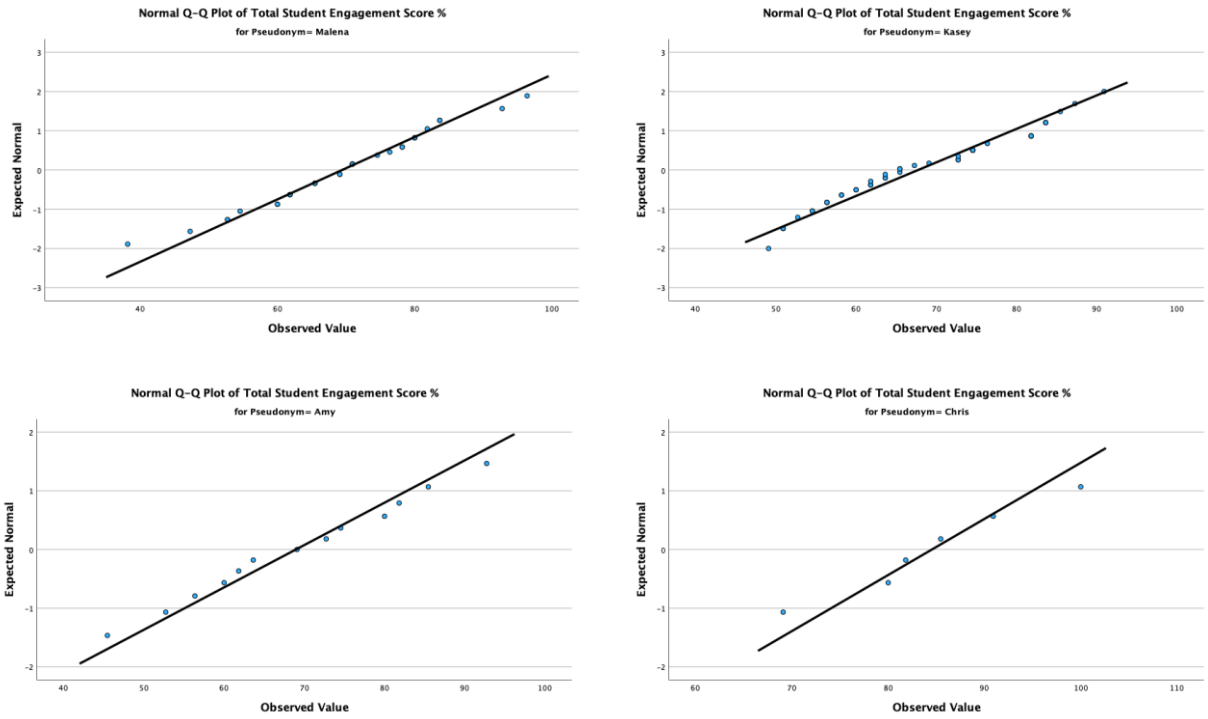
In order to run a one-way ANOVA, there are six assumptions that need to be considered. The first three assumptions relate to the choice of study design and the measurements that has been chosen to make, whilst the second three assumptions relate to how the data fits the one-way ANOVA model.

	Assumption	Test
Assumption #1	One dependent variable (i.e., student engagement and motivation) that is measured at the continuous level.	Passed
Assumption #2	One independent variable that consists of two or more categorical, independent groups (i.e., classroom / teacher).	Passed
Assumption #3	There needs to be independence of observations, which means that there is no relationship between the observations in each group of the independent variable or between the groups themselves.	Passed
Assumption #4	There should be no significant outliers in the groups of your independent variable in terms of the dependent variable.	Passed[a]
Assumption #5	The dependent variable should be approximately normally distributed for each group of the independent variable.	Passed[b]
Assumption #6:	There needs to be homogeneity of variances (i.e., the variance is equal in each group of your independent variable).	Passes[c]

[a] Assumption #4: There were no outliers in the data, as assessed by inspection of a boxplot.



[b] Assumption #5: data was normally distributed for each group, as assessed by visual inspection of Normal Q-Q Plots.



[c] Assumption #6: there was homogeneity of variances, as assessed by Levene's test of

homogeneity of variances (p = .719).

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Total Student Engagement Score %	Based on Mean	.448	3	91	.719
	Based on Median	.395	3	91	.757
	Based on Median and with adjusted df	.395	3	87.467	.757
	Based on trimmed mean	.445	3	91	.721

One-way ANOVA was statistically significant,

Total Student Engagement Score %

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1495.599	3	498.533	3.320	.023
Within Groups	13664.523	91	150.160		
Total	15160.122	94			

Calculating and reporting an effect size

There is more than one method of calculating an effect size for a one-way ANOVA. We selected the method called omega squared (ω2) (Lund & Lund, 2020). This is calculated as:

$$\hat{\omega}^2 = \frac{SS_b - (df_b)MS_w}{SS_t + MS_w} \qquad \omega^2 = \frac{1495.599 - (3)150.160}{15160.122 + 150.160} = 0.068$$

Post Hoc Tests

The one-way ANOVA test cannot tell which specific groups were significantly different from each other; it only tells if a least two groups were different (Lund & Lund, 2020). Since we had four groups in our study, we ran a post hoc test to test all possible group comparisons. Additionally, we decided to run the Tukey-Kramer post hoc test instead of the Tukey test because we had different number of participants (i.e., students) in each group.

Multiple Comparisons

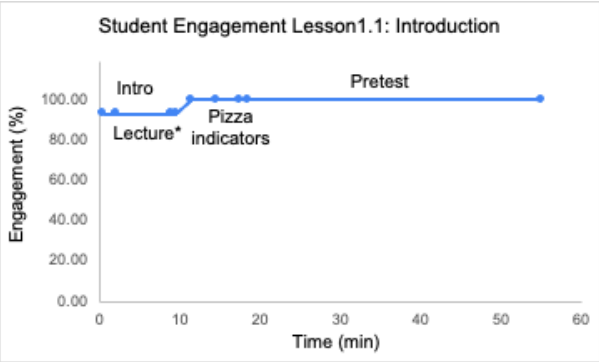
Dependent Variable: Total Student Engagement Score %

	(I) Teacher Pseudonym	(J) Teacher Pseudonym	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	Malena	Kasey	1.68364	2.83591	.934	-5.7383	9.1056
		Amy	.47044	4.01261	.999	-10.0311	10.9720
		Chris	-15.12397*	5.43846	.033	-29.3572	-.8907
	Kasey	Malena	-1.68364	2.83591	.934	-9.1056	5.7383
		Amy	-1.21321	3.87851	.989	-11.3638	8.9374
		Chris	-16.80761*	5.34029	.012	-30.7839	-2.8313
	Amy	Malena	-.47044	4.01261	.999	-10.9720	10.0311
		Kasey	1.21321	3.87851	.989	-8.9374	11.3638
		Chris	-15.59441	6.04792	.055	-31.4227	.2339
	Chris	Malena	15.12397*	5.43846	.033	.8907	29.3572
		Kasey	16.80761*	5.34029	.012	2.8313	30.7839
		Amy	15.59441	6.04792	.055	-.2339	31.4227
Games-Howell	Malena	Kasey	1.68364	2.82450	.933	-5.7599	9.1272
		Amy	.47044	4.42277	1.000	-11.8952	12.8360
		Chris	-15.12397	4.79889	.054	-30.5421	.2941
	Kasey	Malena	-1.68364	2.82450	.933	-9.1272	5.7599
		Amy	-1.21321	4.23719	.992	-13.2220	10.7956
		Chris	-16.80761*	4.62842	.034	-32.2034	-1.4118
	Amy	Malena	-.47044	4.42277	1.000	-12.8360	11.8952
		Kasey	1.21321	4.23719	.992	-10.7956	13.2220
		Chris	-15.59441	5.74503	.074	-32.4814	1.2926
	Chris	Malena	15.12397	4.79889	.054	-.2941	30.5421
		Kasey	16.80761*	4.62842	.034	1.4118	32.2034
		Amy	15.59441	5.74503	.074	-1.2926	32.4814

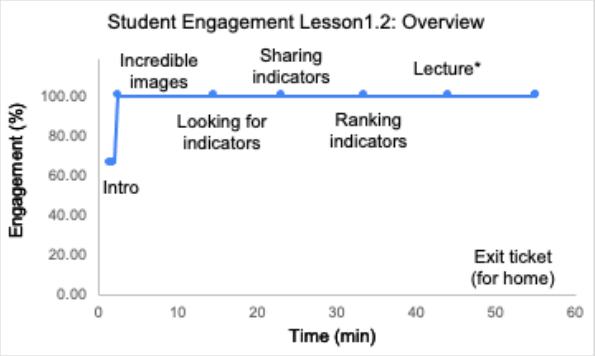
*. The mean difference is significant at the 0.05 level.

Appendix D – Engagement Scores Figures by Case

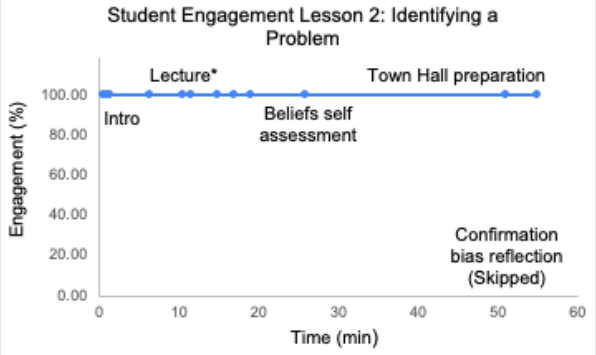
Student Engagement by Activity: KASEY



* Evaluation concepts



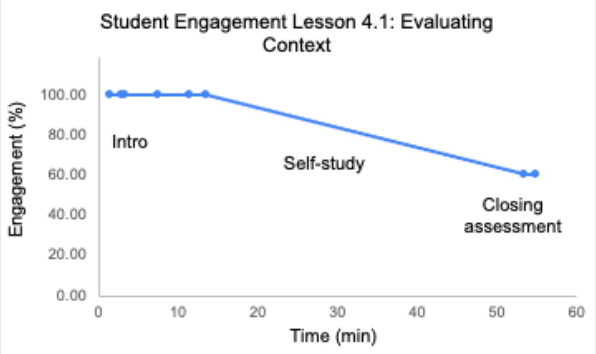
* The Three Tier Framework

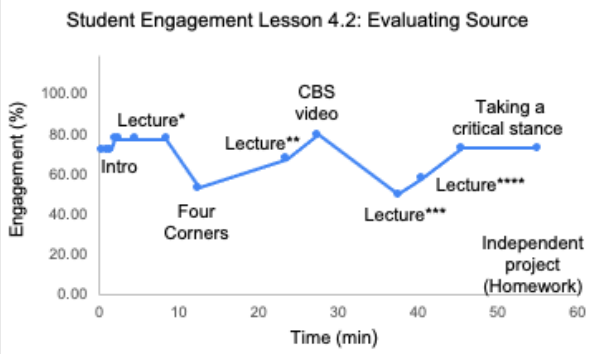


* Perspective, beliefs, prior knowledge

* Web and meta search engines, results page, algorithms, and query syntax

** Filter bubbles, echo chambers, fake news



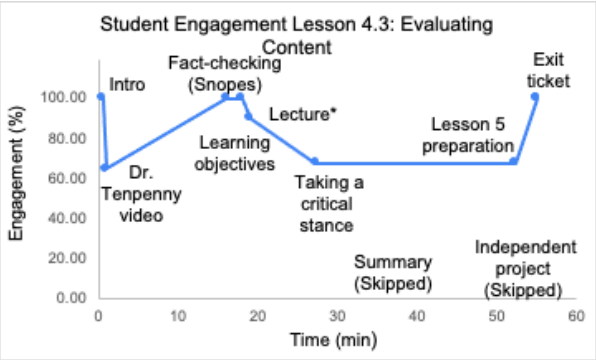


* CORE concepts (evaluation, credibility, relevancy)

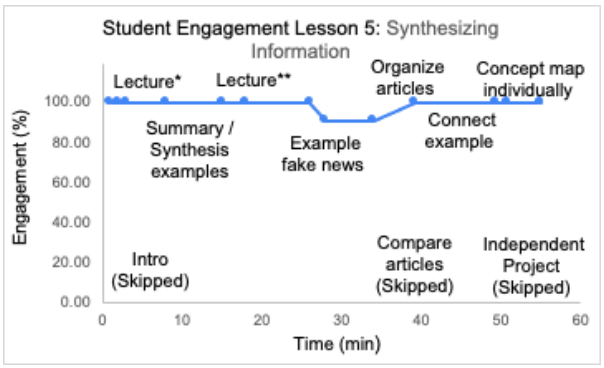
** Types of Organizations

*** Web of Understanding

**** Author's Expertise



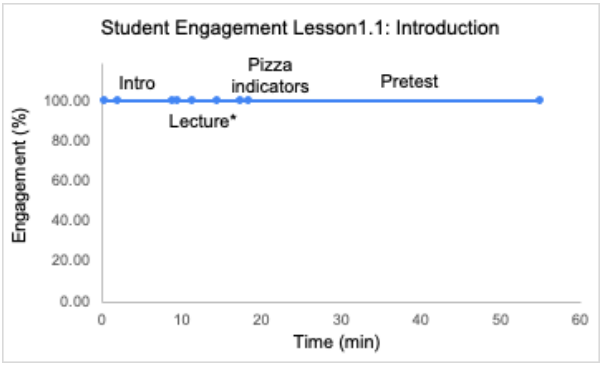
* Content credibility



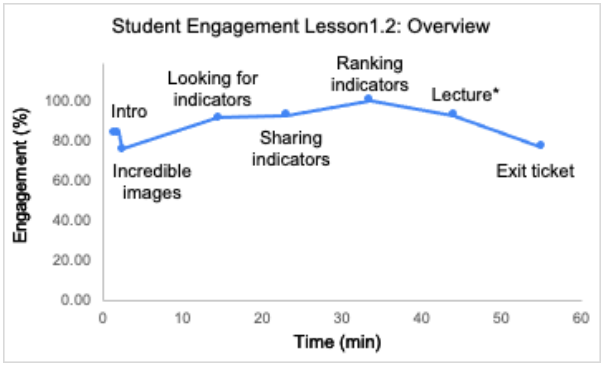
* Summary vs. Synthesis

** Concept map

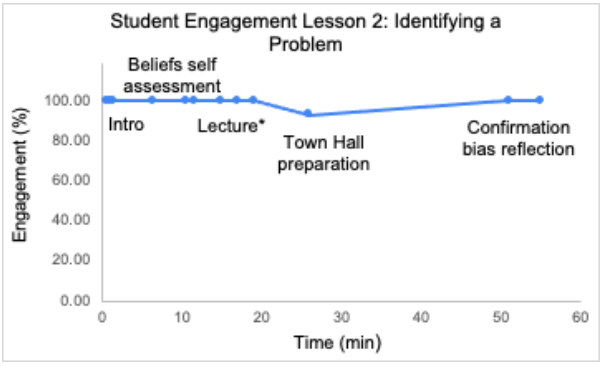
Student Engagement by Activity: MALENA



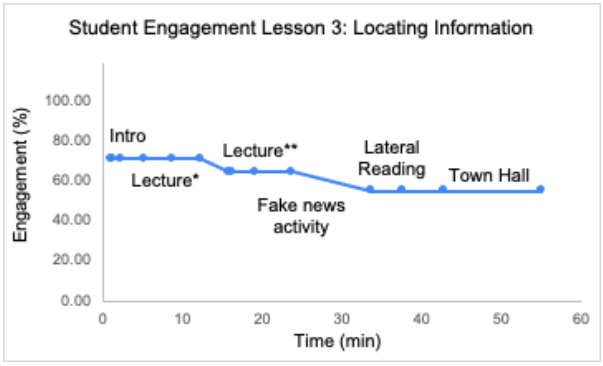
* Evaluation concepts



* The Three Tier Framework

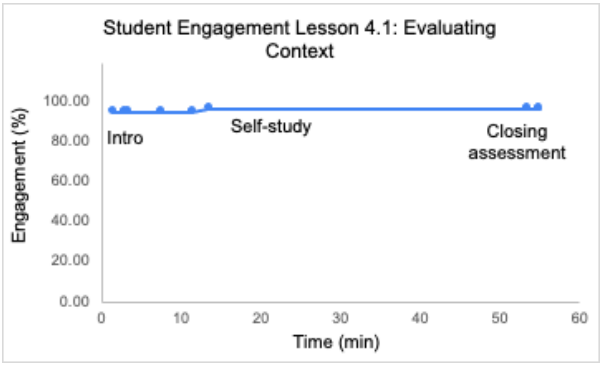


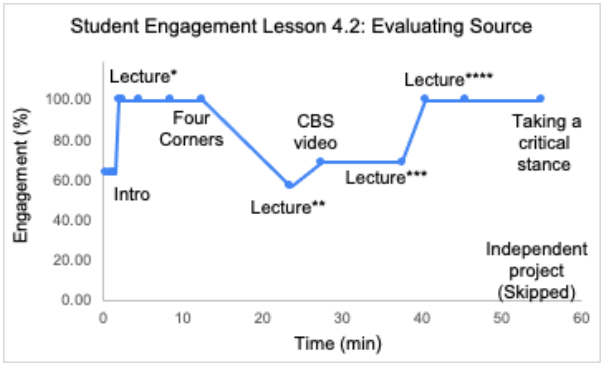
* Perspective, beliefs, prior knowledge



* Web and meta search engines, results page, algorithms, and query syntax

** Filter bubbles, echo chambers, fake news



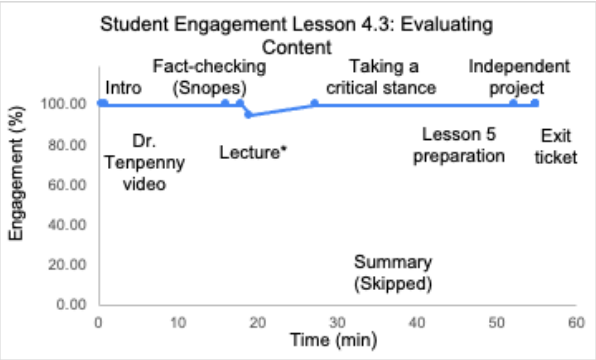


* CORE concepts (evaluation, credibility, relevancy)

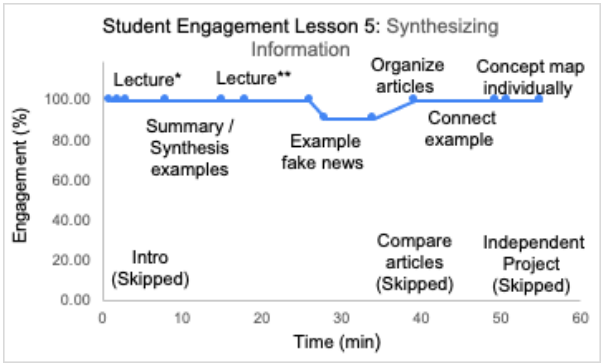
** Types of Organizations

*** Web of Understanding

**** Author's Expertise

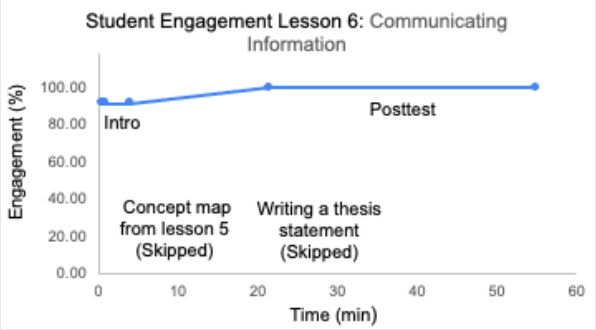


* Content credibility

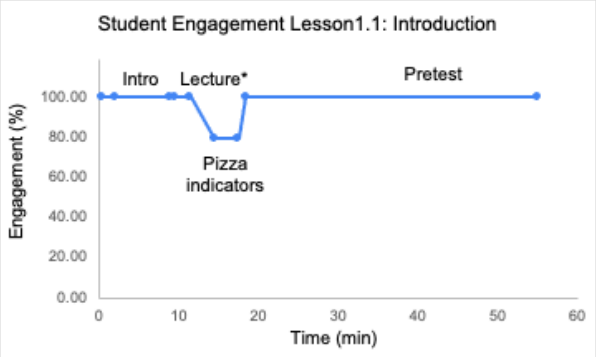


* Summary vs. Synthesis

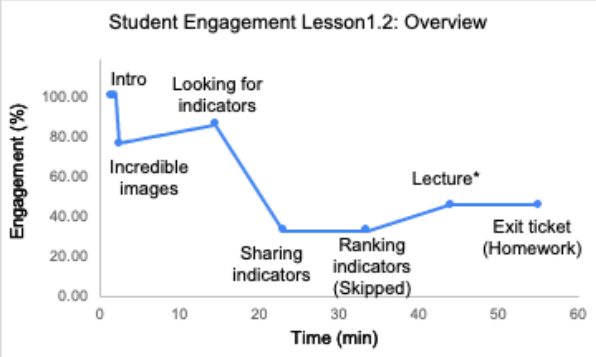
** Concept map



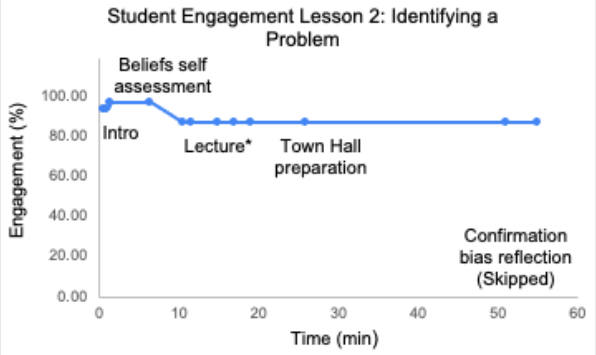
Student Engagement by Activity: AMY (51)



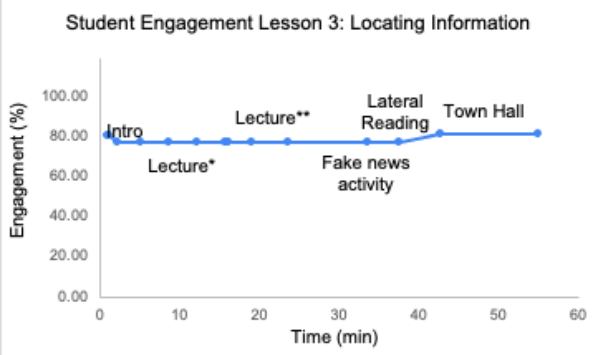
* Evaluation concepts



* The Three Tier Framework

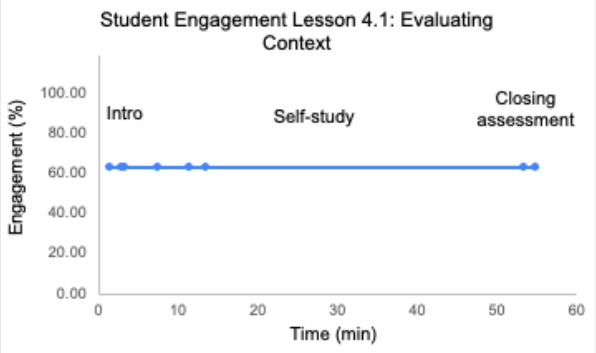


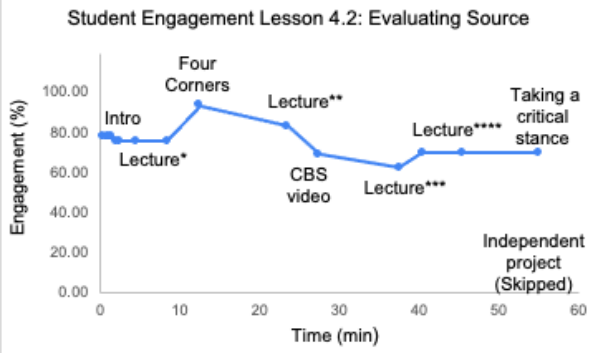
* Perspective, beliefs, prior knowledge



* Web and meta search engines, results page, algorithms, and query syntax

** Filter bubbles, echo chambers, fake news



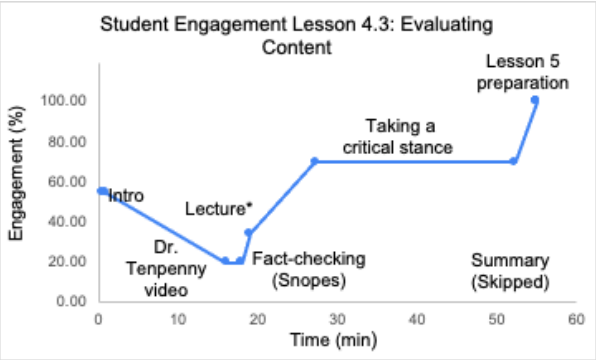


* CORE concepts (evaluation, credibility, relevancy)

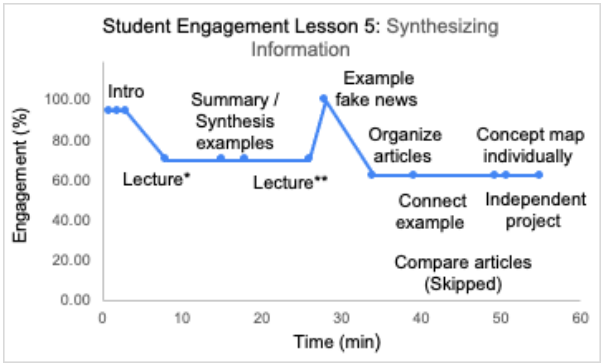
** Types of Organizations

*** Web of Understanding

**** Author's Expertise



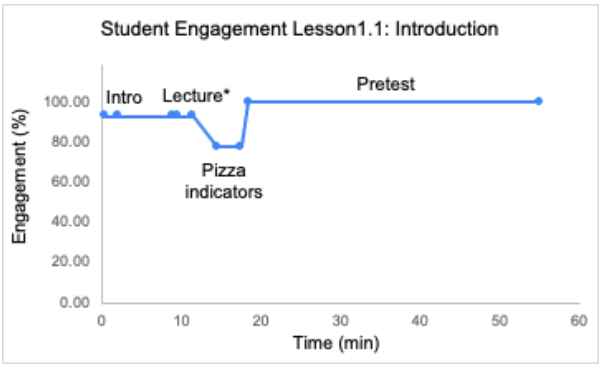
* Content credibility



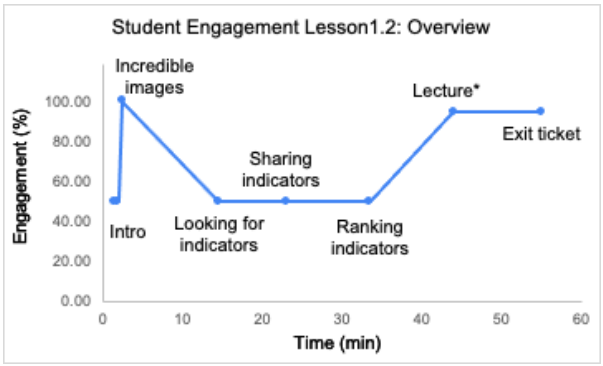
* Summary vs. Synthesis

** Concept map

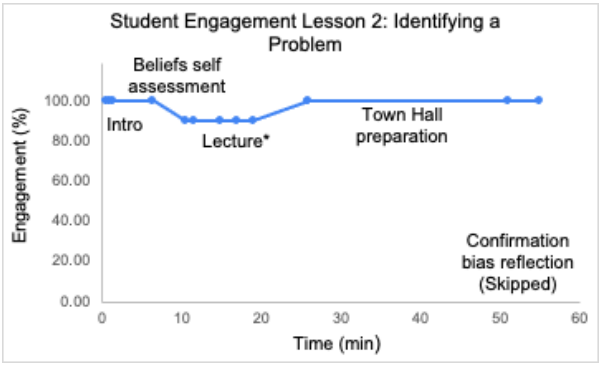
Student Engagement by Activity: AMY (55)



* Evaluation concepts



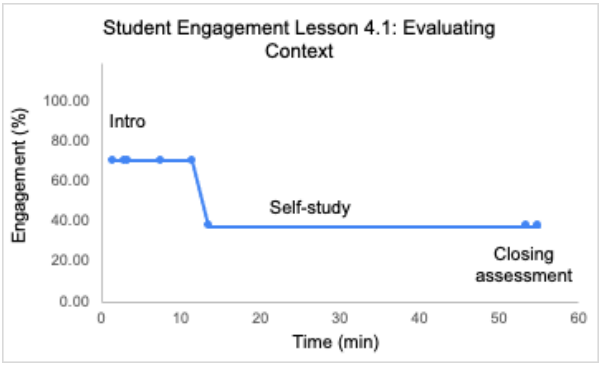
* The Three Tier Framework

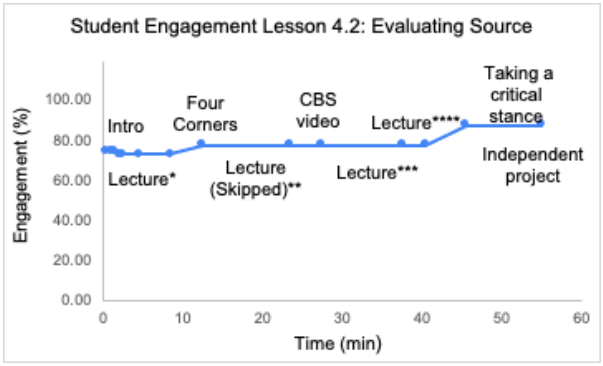


* Perspective, beliefs, prior knowledge

* Web and meta search engines, results page, algorithms, and query syntax

** Filter bubbles, echo chambers, fake news



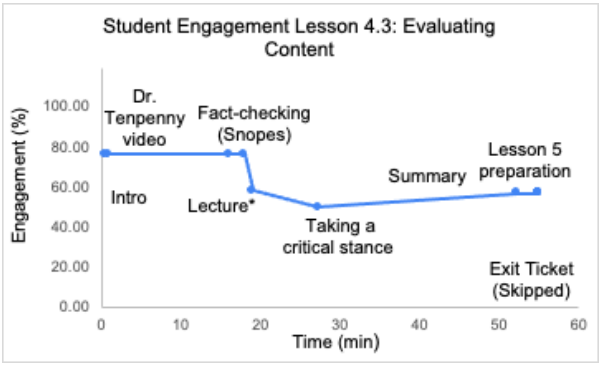


* CORE concepts (evaluation, credibility, relevancy)

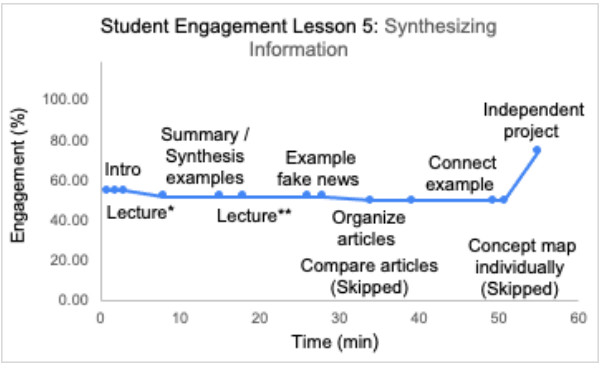
** Types of Organizations

*** Web of Understanding

**** Author's Expertise

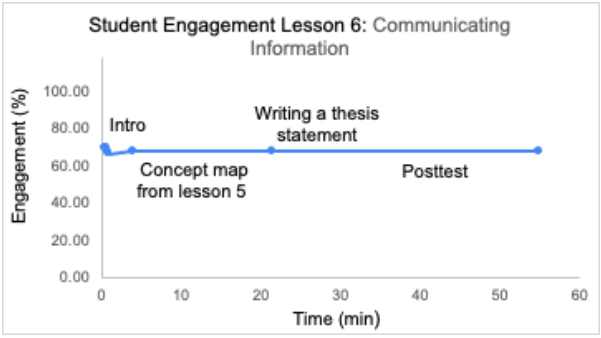


* Content credibility

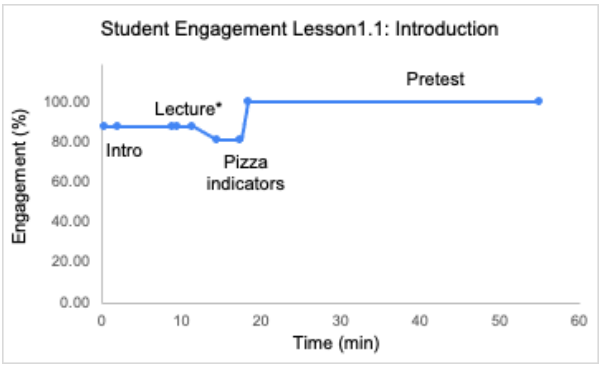


* Summary vs. Synthesis

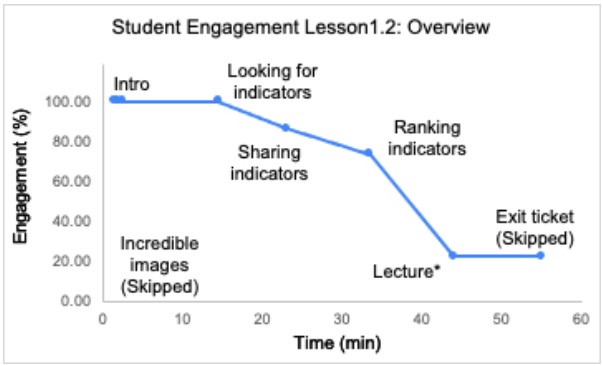
** Concept map



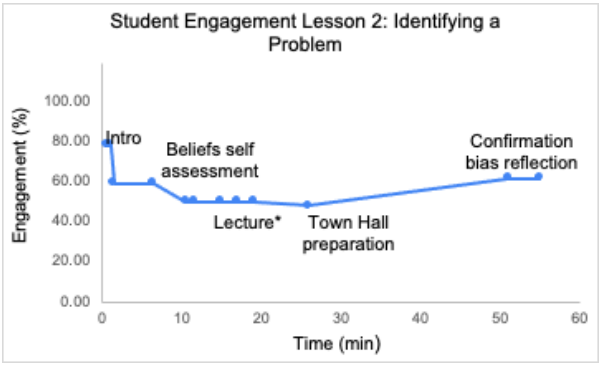
Student Engagement by Activity: CHRIS



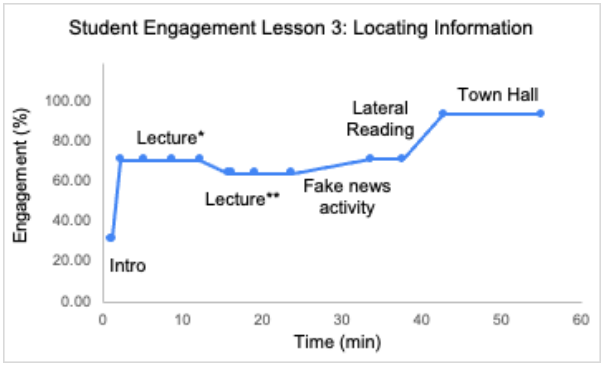
* Evaluation concepts



* The Three Tier Framework

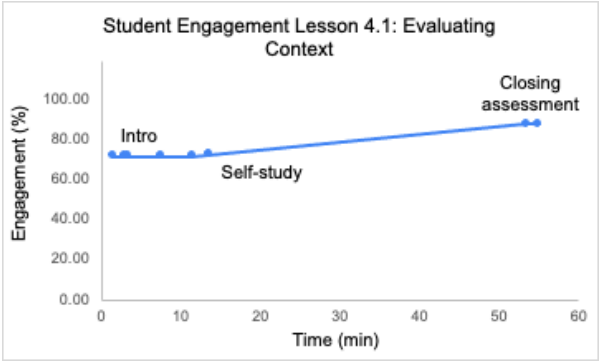


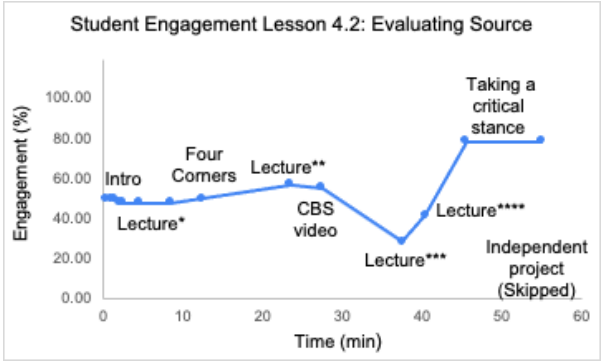
* Perspective, beliefs, prior knowledge



* Web and meta search engines, results page, algorithms, and query syntax

** Filter bubbles, echo chambers, fake news



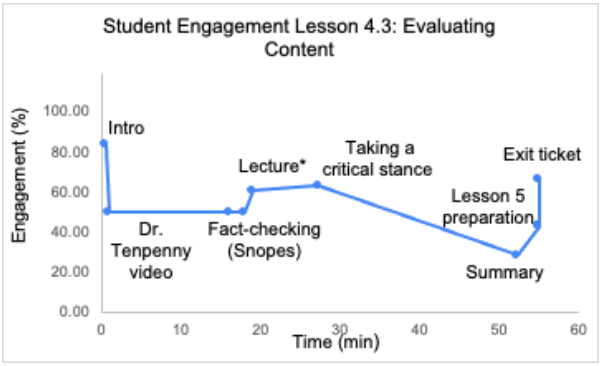


* CORE concepts (evaluation, credibility, relevancy)

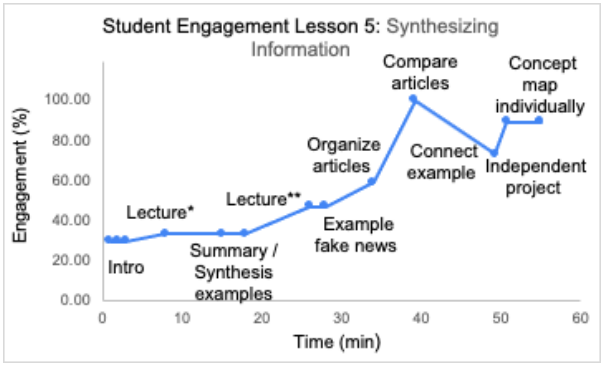
** Types of Organizations

*** Web of Understanding

**** Author's Expertise



* Content credibility



* Summary vs. Synthesis

** Concept map