COGNITIVE STRUCTURES OF DIFFERENTIAL GEOMETRY AND THEIR IMPLICATIONS ON UNDERGRADUATE MATHEMATICS CURRICULUM

By

JULIE JENKINS, B. Sc. H

MASTER OF SCIENCE (2020)
(Mathematics and Statistics)

McMaster University
Hamilton, Ontario

TITLE: Cognitive Structures of Differential Geometry and Their Implications on Undergraduate Mathematics Curriculum

AUTHOR: Julie Jenkins

SUPERVISOR: Dr. Miroslav Lovric

NUMBER OF PAGES: 109

Abstract

Based on similar studies using the FCI and CCI, this project aims to explore how effectively undergraduate courses prepare students for study in higher level mathematics courses. To investigate this, we choose to study the preparedness of Level III differential geometry students by developing and implementing a concept inventory that measures the cognitive structures of prerequisite undergraduate material. Using techniques in item analysis and concept mapping, we assess the cognitive structures of the incoming students, and identify areas for improvement within the Calculus and Linear Algebra course sequences based on the current literature on concept inventories. We also investigate potential relationships between cognitive structure and academic success, and attempt to measure the development of cognitive structures as a result of instruction throughout the term.

Acknowledgements

This thesis marks the end of my time at McMaster. I'd like to thank Department of Mathematics and Statistics, which is full of professors that went out of their way to see me succeed. This department holds a very special place in my heart; I am so grateful to have called it home.

I would like to thank Dr. Lovric, for being the best supervisor I could have asked for. Thank you for believing me, even when I didn't, and for actively creating opportunities for me to do what I love. You are my biggest advocate and my most trusted mentor. I am truly thankful to have you in my life.

Thank you to Dr. Nicas, for supporting this project, looking out for my best interests, and encouraging me to try new things. Your faith in me has made me a braver, more adventurous student. I look forward to our continued friendship, and very much hope to make you proud.

Many thanks to my friends—Genevieve Maalouf, Dicle Mutlu and Gregory Forkutza—for their support and friendship over the past two years. Thank you to Fletcher Gates, for being the best editor in chief, and my favorite person to debate educational philosophies with. Of course, I owe a very heartfelt thank you to Cam, who lifts me up like no one else can.

Lastly, I need to thank my family, for being an endless source of encouragement. Not a day goes by when I don't think about how lucky I am to be a Jenkins, with an army of energetic, kind-hearted and compassionate people behind me. I thank my sisters, who inspire me constantly and treat me with the upmost empathy. Most importantly, I thank my parents, for doing everything they could to make sure I could go to school, despite all the setbacks that came our way. Mom and Dad—we made it. It is an honour to share this with you.

Table of Contents

List of Figures 6
Introduction: 7
Literature Review: 9
Research Methods: 16
Results: Phase I 31
Results from Phase II: 40
Discussion and Analysis: 50
Conclusion: 62
Bibliography 64
Appendix A: Sample Validity Tests from Cognitive Labs 66
Appendix B: Sample Peer Evaluation 89
Appendix C: Finalized Item Coverage for Inventories 2 and 3 92
Appendix D: Items in Categories (1) and (3) 93
Appendix E: Letter of Information, Phase I 97
Appendix F: Oral Consent Script, Phase I 100
Appendix G: Online Survey Screening Questions, Phase I 102
Appendix H: Recruitment Email On behalf of Researcher, Phase I 103
Appendix I: Letter of Information, Phase II 104
Appendix J: Recruitment Script for Phase II 108
Appendix K: Recruitment Email On behalf of Researcher, Phase II 109

List of Figures

Figure 1 Generating student concept maps based on concept inventory responses. Figure taken from Castles and Lohani. 14
Figure 2 Overview of Research Plan. 17
Figure 3: Overview of methodology for creating the first instrument draft 18
Figure 4 Sample question Analysis. Under "Content Validity", we considered Notation [N], Terminology [T], and Confusing wording [C] validity in accordance with the criticisms of the CCI presented in Gleason, et. Al, 2015. 20
Figure 5 Overview of validity testing in Phase I. 21
Figure 6: Translation from sample items to their representation on a concept map 26
Figure 7: Framework for interpreting the difficulty index p*, adapted from (Bai & Ola, 2017).* 27
Figure 8: Framework for interpreting the discrimination index, adapted from (Bai & Ola, 2017) 28
Figure 9: Interpretive framework for difficulty and discrimination indices used in analyzing item responses from inventories. 29
Figure 10: Item topic coverage for prerequisite concept inventory. 32
Figure 11 Summary of references to the past and computation with respect to final score and items attempted 36
Figure 12 Sample Assumption Notation 37
Figure 13: Base Concept Map for Pre-Requisite Concept Inventory. Available online here, using the URL: https://lucid.app/lucidchart/15032623-43cc-4084-b923-a4202344e406/view 39
Figure 14 Gender distribution, $n = 108$*.* 40
Figure 15 Citizenship distribution of the sample, $n = 101$ 41
Figure 16 Age distribution as indicated by partial birthdate, $n = 108$*.* 41
Figure 17 Summary of item analysis from Inventory 1, based on 110 valid responses. 42
Figure 18: Interpretation framework for items in inventory 1. 43
*Figure 19 Distractor Summary for items in categories (1) and (3). Note that "total responses" here refers to the number of responses that come from valid inventory submissions; that is, responses from participants who answered at least 19 of the 22 items. The correct response is indicated by *.* 44
Figure 20 Quartile 2 map for inventory 1, using a cut off of 55 correct answers. Available online here, using the URL https://lucid.app/lucidchart/15032623-43cc-4084-b923-a4202344e406/view 46
Figure 21 Quartile 3 map for inventory 1, using a cut off score of 83. Available online here, using the URL: https://lucid.app/lucidchart/15032623-43cc-4084-b923-a4202344e406/view 47
Figure 22 Summary of item analysis for Inventory 2, based on 96 valid responses 48
Figure 23 Summary of item analysis for inventory 3, based on 89 valid responses 48

Introduction:

Among the many challenges of undergraduate teaching is the instructor's inability to predict, with some certainty, what a new cohort of students will understand when entering their courses. Students who have enrolled in upper-level undergraduate classes have, at minimum, taken similar prerequisite courses. In theory, students acquire the necessary knowledge and skills that prepare them adequately for further advanced courses by passing through a prerequisite course sequence.

However, establishing what students *actually* understand is extremely difficult, even when prerequisite sequences are completed at same institution. This is especially true in mathematics, where large class sizes and breadth of content required to cover impose substantial limitations on assessments. Though traditional assessments do not typically test computational or routine applications exclusively, the limited resources available for administering and grading divergent conceptual tasks limit their use in mathematics courses. This means that students can compute their way through course content, without having developed or demonstrated deep understanding of the material. While institutions prevent students from enrolling in courses for which they have not received prerequisite credits, instructors have no guarantee that students who do enter their classes will have the prerequisite knowledge necessary to succeed in their courses.

Early studies that address the issue of students' conceptual understanding in physics found that students entering their undergraduate classes have beliefs about the subject and the content that directly contradict the material that they learn in class, and more surprisingly, that the misconceptions they have when entering their courses are typically not eliminated as a result of the course instruction. This discovery revolutionized physics instruction and research in undergraduate physics education, resulting in the development of test instruments (such as the Force Concept Inventory) that effectively measure conceptual understanding. As well, instructional strategies which favour interactive engagement learning models have replaced traditional lecture styles, partially or almost completely.

Similar studies have been conducted in undergraduate mathematics education and echo the findings of the physics educators; students enter university with misconstrued understandings of mathematical ideas and carry them into their later coursework regardless of further instruction in the subject. Though there have been substantial breakthroughs in this area for students entering undergraduate studies in mathematics, the existing literature on student understanding of material taught in first- and second- year mathematics courses is scarce. In essence, we don't know the conceptualizations of mathematics students bring into their mathematics courses.

This reality motivates the first goal of our study, which is to better understand the conceptual understanding of students as they conclude their prerequisite course sequences and embark on their first upper-level mathematics courses. In doing so, we are able to evaluate the prerequisite course sequences in terms of their ability support the development of student understanding and identify areas in the curriculum where students appear to experience difficulties.

To achieve this goal, we explore the position of MATH 3B03, a third-year differential geometry course at McMaster University, in the undergraduate mathematics curriculum. We choose this course because the prerequisite courses for MATH 3B03 include most required courses in the calculus and linear algebra sequence for the Bachelor of Science programs in mathematics, and because the subject matter of the course aligns with the content taught in previous courses. This makes students in the course an ideal sample to consider when evaluating the prerequisite courses' ability to prepare students for their studies in upper-level mathematics. In particular, our first research question is:

> To what extent are the prerequisite mathematics courses appropriate in terms of building conceptual understanding necessary for students to succeed in higher level mathematics courses, such as MATH 3B03?

The second goal of our study is to determine whether students' conceptual framework for mathematics would positively correlate to academic success. The literature is divided on the relationship between a students' cognitive structure of a discipline (that is, their conceptual framework of the discipline) and academic success in this area, as academic success and conceptual understanding are not identical variables. However, theory asserts that a student's cognitive structure will be a determinant of academic success in this context, since upper-level mathematics courses distinguish themselves from earlier mathematics courses based on the requirement for students to integrate a multitude of concepts to solve problems and understand and work on proofs. This motivates the second research question for our investigation:

> Is there a measurable relationship between cognitive structure and academic success, and if so, how can we support the development of rich cognitive structures for students within the McMaster undergraduate mathematics curriculum?

During our study, we stumbled into an unexpected barrier. We needed an instrument that can evaluate students' cognitive structure of prerequisite material for differential geometry but realized that no such instrument existed. Consequently, the focus and scope of our research expanded to include the development and validation of appropriate test instruments for the study. We address the research questions above using data collected and verified founded on the concept inventories that we developed, and model student cognitive structure using students' responses to the inventory instruments.

Literature Review:

The body of existing research into conceptual understanding of mathematics is focused on primary and secondary education, with some studies focusing on the transition between secondary and post-secondary education. In undergraduate level mathematics, it can be challenging to construct questions that explicitly measure conceptual understanding in isolation of procedural or computational knowledge in the first two years of study, due to limited resources for grading divergent tasks. In practice, measures of student understanding in post-secondary education include tests, exams and assignments, for which students obtain a grade point for their performance in a course. However, it is unclear what level of understanding separates two students that achieve the same or similar grade levels (Castles & Lohani, 2009) .

Concept Inventories:

An alternative to using term-grades as a metric of conceptual understanding is the use of concept inventories. These specially designed assessments aim to evaluate a student's conceptual understanding of material (Furrow & Hsu, 2019). Concept inventories are typically multiple-choice instruments, but there are a few concept inventories that are exclusively open-ended or that are a mix of open-ended and multiple-choice items for upper-level courses (Madsen, McKagan, & Sayre, 2017).

While the current literature is limited in the context of post-secondary mathematics education, there have been a wealth of studies that examine the use of concept inventories in other STEM related fields. In these disciplines, concept inventories have been shown to have logistical advantages as a measure of conceptual understanding. As concept inventories are multiple choice, they require very little effort in administer and grade (Furrow & Hsu, 2019). Further, effective concept inventories are composed of questions that target common misconceptions and essential concepts for students, making them invaluable resources for guiding instruction and alleviating student misconceptions (Furrow & Hsu, 2019).

Concept inventories have been particularly useful in identifying challenges for students and measuring the effect that a semester of instruction has on addressing those challenges. Through the development of the Force Concept Inventory (FCI) in physics, researchers discovered that many students had developed conceptions of physical ideas and concepts that directly contradicted the curriculum that they were learning (Epstein, 2013). Findings using the FCI have revolutionized undergraduate education, the most provocative of being that a semester of classes made very little difference to alter misconceptions have in physics (Epstein, 2013). This is especially critical when considering that a students' performance on the FCI has been shown to have higher predictive validity of a students' academic success within a course than any other indicator, including age, gender, and academic background (Epstein, 2013). After shifting instructional methods employed in physics classes from traditional lecture style to those that use interactive engagement, students showed higher gains on the FCI, informing the development of new resources for physics instruction and teaching methods (Epstein, 2013).

After observing the effects of the FCI in physics, there have been some attempts to create a similar model for post-secondary mathematics through development of a Calculus Concept Inventory, or CCI (Epstein, 2013). Developed by Jerome Epstein between 2006 and 2008, the CCI aims to measure students conceptual understanding when they arrive in their first university mathematics calculus courses (Epstein, 2013). The concepts tested come from secondary mathematics curricula, and include material that, according to Epstein, are fundamental and universal concepts that faculty assume students have a strong grasp of coming into post-secondary calculus (Epstein, 2013). Thus far, the results of the CCI have mirrored those of the FCI; students have deeply misconstrued perceptions of the underpinning mathematical concepts that are required to succeed in post-secondary mathematics courses (Epstein, 2013). Instructors saw higher normalized gains in the CCI results and higher academic performance among calculus students in classes that favour interactive engagement strategies when compared to their peers who are taught using traditional lecture style classes (Epstein, 2013).

The ability of the CCI to diagnose student misconceptions and to measure the change in conceptual learning over a semester has cause it to be evaluated as a new metric for inform educational reform in post-secondary mathematics. The inventory has enabled studies that investigate the use of interactive engagement strategies and concept inventories as formative assessments in mathematics (Lai, 2009). Related projects that have emerged include the *Good Questions* project developed at Cornell University, and the Basic Skills Concept Inventory, which are used to probe for misconceptions via active learning and diagnostics, respectively (Lai, 2009). Themes from these initiatives include making key concepts in introductory calculus clearly visible for students and providing them with opportunities to identify areas of weakness through formative feedback (Lai, 2009).

Though the CCI and Basic Skills Inventory offer valuable insight on incoming Level I students, there has yet to be a concept inventory developed for students after their preliminary course sequences in mathematics. The CCI has an extremely narrow focus on the type of understanding it intends to assess, which makes it inappropriate for upper-level students. Furthermore, the concept inventory as it is currently developed for calculus does not provide insight as to the general structure of a students' understanding of the material, making the results of the inventory challenging to interpret from an instructional point of view.

Concept inventories have not been widely adopted by post-secondary mathematics instructors. This is partially due to variances in disciplinary coverage, limited evidence of validity and reliability of concept inventories in their fields and understanding how to effectively use concept inventories (Furrow & Hsu, 2019). It has been proposed that interpreting concept inventory scores is difficult, limiting their use in educational settings (Furrow & Hsu, 2019). Anecdotal evidence suggests that many instructors are not aware of concept inventories, their purposes or their uses.

As a result of the limited coverage and logistical challenges associated with adopting concept inventories in higher level mathematics, there is limited research as to what students beyond Level I and II truly understand about mathematics that uses validated concept inventories. This includes research on how effective prerequisite course sequences are in preparing students for studies in higher level mathematics. The lack of instrumentation in mathematics education imposes additional barriers to new educational research on subjects beyond first- and second-year calculus sequences. For instance, we were unable to find a concept inventory that would suitably measure concepts needed for an introductory differential geometry class.

Our work aims to address these challenges by constructing a mathematics concept inventory that is relevant for students halfway through their undergraduate mathematics education. We will also provide evidence of validity for the pilot study and including suggested uses for instructors should they choose to adopt the inventory for themselves. By doing so, we hope to leverage the logistical and cognitive advantages of using concept inventories, while mitigating some of the barriers to implementation.

Concept Maps:

Another alternative to traditional assessments of understanding are concept maps. These are graphic representations of how a curriculum organizes, or a student understands, concepts with respect to other concepts. Much like concept inventories, concept maps are intended to assess conceptual understanding; however, concept maps have the added advantage of providing a clear representation of student knowledge, making them easily understood by practitioners.

Developed from J. Novak's research in meaningful learning, concept maps are intended to summarize and organize understanding of a subject (Institute for Machine and Human Cognition, 2003). Within this theory, the term *"concept"* is defined to be "*a perceived regularity in events or objects, or a record of events or objects, designated by a label"* (Institute for Machine and Human Cognition, 2003). Within a concept map, concepts are represented by nodes, and are often labeled with one or two words, such as "derivative" or "monotone"; any statement about or involving the concept is defined as a *proposition*, denoted by an arrow that connects two nodes, with a label that describes the relationship between conceptual objects (Institute for Machine and Human Cognition, 2003). The resulting concept map is a graphical representation of a student's *cognitive structure*, defined as "*hierarchically organized in terms of highly inclusive concepts under which are subsumed less inclusive sub concepts and informal data"* (Ivie, 1998).

Novak's representations of cognitive structure stem from Ausubel's Theory of Assimilation, which emphasizes the integration of new knowledge within an existing cognitive structure (Ivie, 1998). The model assumes an existence of a cognitive structure within a student's thinking process, that is organized hierarchically, with large, broad concepts at the top, and increasingly more particular concepts organized below (Ausubel, 1963). The theory implies that learning occurs when new concepts are correctly and clearly anchored to an appropriate concept based on the relationship between concepts (Ausubel, 1963). If a concept is anchored to relevant and appropriately general concepts with a high degree of organization, then the learning is said to be *meaningful;* however, if the ideational structure is ill-prepared to subsume a concept (e.g., due to being disorderly, too general, etc.), the resultant cognitive structure remains unstable and are more likely to be forgotten, or unable to be used in abstract problem solving (Ausubel, 1963).

Though concept maps are representations of cognitive structures, which are presumed to be highly organized, it is important to note that experts in a field will often produce concept maps that appear chaotic (Madsen, McKagan, & Sayre, 2017). This is because experts have more experience with material and are able to form more conceptual connections between seemingly distinct concepts, referred to as "crosslinks" (Madsen, McKagan, & Sayre, 2017). The presence of accurate cross links is evidence of expert level knowledge, which can make the maps of experts look comparatively less clean than a novice map, despite being highly organized.

Ausubel's Theory of Assimilation is a reasonable theoretical framework for mathematics due to the structural nature of mathematics as a discipline; since new concepts in mathematics rely on those previously established, the nature of mathematical learning aligns well with a propositional model of knowledge assimilation. Additionally, the validity demonstrated by concept maps as representations of cognitive structures, we will use concept maps as models for conceptual understanding of mathematics to communicate results of the concept inventory we develop.

Much like concept inventories, concept maps have not been extensively studied in mathematics education but have been examined as measures of conceptual understanding in other STEM fields. For instance, Novak's original interest in concept mapping was inspired by teaching elementary level science concepts (Institute for Machine and Human Cognition, 2003). At the post-secondary level, the validity of concept maps as a divergent psychometric task has been established in biology majors (Mintzes, Markham, & Jones, 1994). As a general psychometric tool, concept maps and their corresponding scores are related to academic assessments, showing that they measure a related construct to academic performance—albeit not entirely identical to academic performance, since conceptual understanding is often a subcomponent of academic assessments (Institute for Machine and Human Cognition, 2003).

The literature surrounding concept maps as psychometric assessments has primarily focused on student-generated concept maps. Some instructors have given students names of concepts and asked them to link them using propositions of their choice (Institute for Machine and Human Cognition, 2003). Others have given guiding questions and allowed students to construct all their own propositions and concepts for their concept maps (Mintzes, Markham, & Jones, 1994). While both of these methodologies have their pedagogical advantages, it makes scoring and evaluating concept maps time consuming and subjective. It also means that the validity of concept maps depends on the reliability of scores when assessed by a variety of evaluators (Institute for Machine and Human Cognition, 2003). These processes make concept maps difficult to adopt and evaluate.

To address these issues, Furrow and Hsu have suggested creating an automated concept map that is generated in response to a students' concept inventory responses. They suggest generating an expert level map, with each proposition and conceptual node being linked to items in concept inventories that intend to assess (Castles & Lohani, 2009). The concept map would generate automatically once concept maps are scored and present an overview of a students' cognitive structure at the time of concept inventory was administered (Castles & Lohani, 2009).

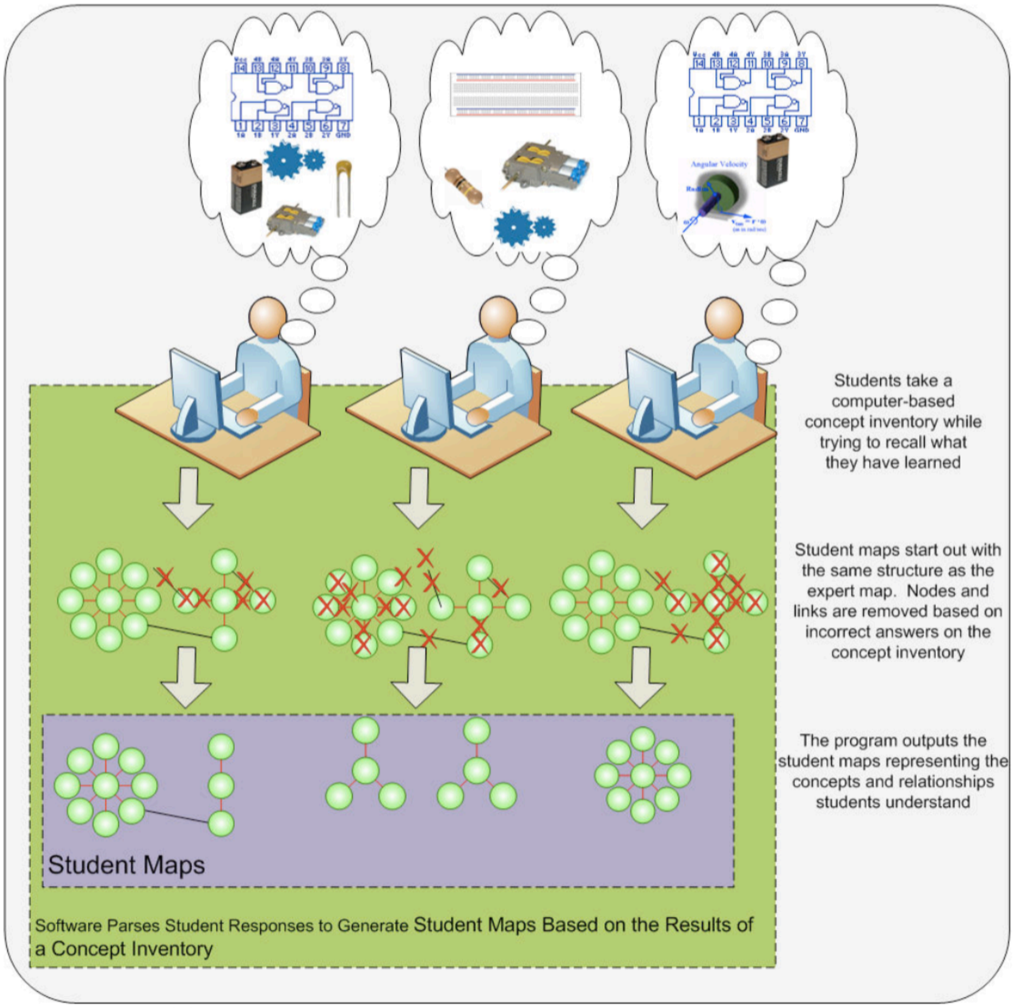


Figure 1 Generating student concept maps based on concept inventory responses. Figure taken from Castles and Lohani.

Castles and Lohani outline a method to develop an auto-generated concept map. They suggest developing an expert-level map, that models all conceptual links that are necessary for a given concept and constructing a concept inventory that has items associated with each proposition (Castles & Lohani, 2009); however, there is no platform currently available that automates the creation of a concept map from a concept inventory. For our project, we will follow a similar methodology to the GCI and CCI's development to construct items. We will extend this inventory into a concept map as suggested by constructing a protype for such an instrument and implement it at the class-level. This will enable us to comment on the cohort's understanding of fundamental course concepts for the purpose of evaluating the prerequisite curriculum.

The suggested tool would preserve the graphical representation of cognitive structure, while leveraging the objectivity and automation aspects of concept inventories. This has not been created as of yet, but we develop and implement a prototype of this instrument and use it to construct a representation of the cognitive structure measured through the concept inventory. Castles and Lohani (2009) suggest creating a map for each student; in our research, we will consider an aggregate map to assess the class as a whole to better comment on the readiness of students for higher-level mathematics instead of on an individual student basis.

Development of Concept Inventories:

The development of a concept inventory takes place in three phases: domain analysis, creation and validation. The CCI, for instance, was developed by a panel of expert faculty that established the basic prerequisite requirements for students, in order for them to have a functional understanding of calculus (Epstein, 2013). Similarly, the first phase of development for a Group Concept Inventory (GCI) was an in-depth literature review and concept domain analysis to construct a taxonomy for group theory, since such a taxonomy did not yet exist (Melhuish, 2015).

Inventory items are created following concept domain analysis, based on concepts that were deemed fundamental. While constructing questions, developers consult a panel of experts to review the items (Melhuish, 2015), and modify them in response to feedback. Once inventories are created, items are validated or modified before being implemented into a larger and more representative sample, typically by performing cognitive labs or giving students free-response versions of test items. Epstein used cognitive labs in the development of the CCI, which are highly structured subject-interviews that aim to uncover how a student processes and reasons through problems (Epstein, 2013), whereas Melhuish (2015) gave modified test-items in an open-response format. Items are then refined, before being tested in a larger sample.

After being tested in a larger, representative sample of students, inventories can be further revised using statistical analysis of student responses (Melhuish, 2015). These techniques include item analysis, distractor analysis and factor analysis, which help developers to revise or eliminate items in the instrument (Jorian, Gane, DiBello, & Pellegrino, 2015). They can also be evaluated in terms of internal consistency through using $KR20$ and $KR21$ coefficients (Frankel & Wallen, 2009). These measures establish the extent to which a concept inventory measures a particular construct; high $KR21$ and $KR20$ coefficients imply that the items in the inventory are highly related to each other, whereas the other measures stated here allow for analysis of each item's validity.

Research Methods:

The goal of this project is to better understand how undergraduates comprehend material taught to them in their early undergraduate courses. If we are successful, we will be able to comment on the effectiveness on these courses, identify areas where our undergraduate students commonly experience difficulty, and propose instructional and curricular recommendations to improve the quality of undergraduate education in mathematics at McMaster.

To investigate this, we examine the place of MATH 3B03 in the undergraduate mathematics program because its' prerequisite course sequence and newly taught content make it uniquely suited to study these structures. As previously stated, we frame our investigation with the goal of answering the following two research questions:

1. To what extent are the prerequisite mathematics courses appropriate in terms of building conceptual understanding necessary for students to succeed in higher level mathematics courses, such as MATH 3B03?
2. Is there a measurable relationship between cognitive structure and academic success, and if so, how can we support the development of rich cognitive structures for students within the McMaster undergraduate mathematics curriculum?

As discussed in the literature review, we were surprised to find that there were no existing instruments that would effectively measure the cognitive structures necessary for higher level maths, which we needed to assess in order to answer our research questions.

Consequently, an additional goal of this project is to construct a concept inventory that would suitably assess the cognitive structures of Level III undergraduate students taking a differential geometry course. Bearing that in mind, our research goals were extended to include the design and validation of a test instrument that would support our original research goals. Using this inventory, we could begin to conjecture as to how well students were prepared for advanced study in mathematics, suggest further areas of investigation, and propose academic support that has been effective in other disciplines in terms of developing conceptual understanding.

Our research was completed in two distinct phases, an instrumentation phase and an implementation phase, which are referred to as "Phase I" and "Phase II", respectively. In our first phase of research, we construct a concept map that encompasses the prerequisite knowledge required to facilitate meaningful learning in differential geometry. We also developed two more concept inventories to measure the development of cognitive structures throughout the term. As in the development of CCI and GCI, we performed cognitive labs and peer review of the instruments constructed to check for issues with validity prior to implementing the instruments in a classroom.

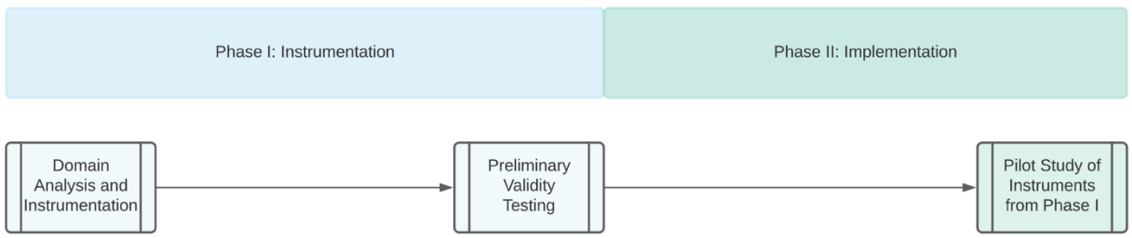


Figure 2 Overview of Research Plan.

The second phase of research was a pilot study, where we test the instruments on a class of MATH 3B03 students over the course of a semester. We administered the pre-requisite inventory early in the term, and the other two in-course inventories prior to term tests. Data collected throughout this phase of research included inventory data, subject-information data and copies of term tests.

Phase I: Instrumentation

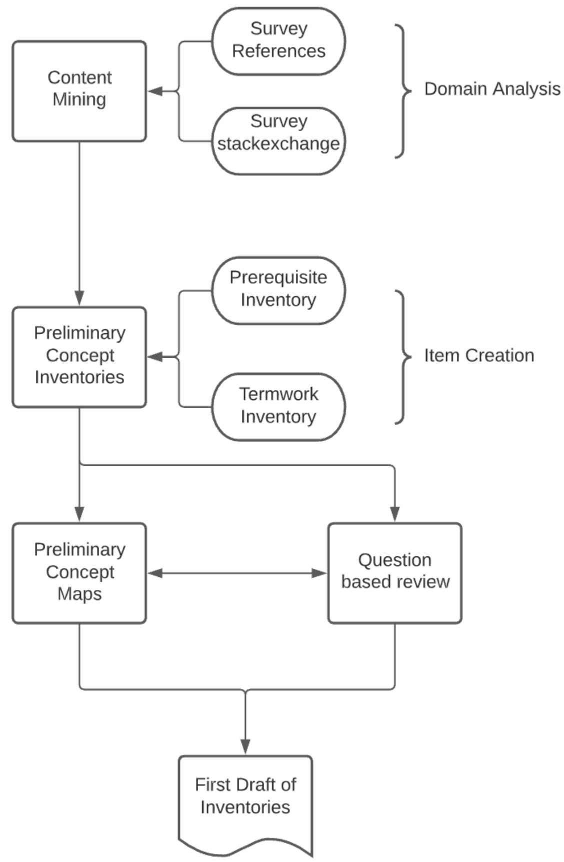


Figure 3: Overview of methodology for creating the first instrument draft

We surveyed the two textbooks used in the course, Presley's *Elementary Differential Geometry*, and Raussen's *Elementary Differential Geometry: Curves and Surfaces*. When examining the texts, we looked for any reference to prerequisite material that was necessary to build a conceptual framework of new proofs, objects or results. For instance, in defining arclength of a curve, we deemed it necessary for students to have a strong understanding of Riemann sums, the length of a vector formula, and tangent vectors, since arclength is defined as the limit of the Riemann Sum of the lengths of secant vectors along the points on a curve. Without prerequisite understanding of these concepts, students may conceptualize arclength as a strictly computational object, rather than truly understand the geometric derivation.

After reviewing the textbooks used for the course, we consulted the *Good Questions* project, which is a question bank of multiple-choice items intended to be used in lecture to identify student misconceptions through active learning (Terrell, n.d.). These questions target conceptual issues and have associated commentary throughout the question bank that discuss possible reasons why students have chosen certain distractor items[1] in the past. On occasion, we adapted questions from this bank, since they were field-tested items that targeted the misconceptions that were relevant to our investigation.

Reading through student-led discussion forums to guide item creation is our unique approach the development of this instrument. To complete our domain analysis, we combed through StackExchange to see what students commonly identified as challenging within their differential geometry practice problems. We searched for articles on any question identified within the textbooks for chapters being covered throughout the course and looked for any reference to particular difficulties students had within their question or within the discussion threads. We also examined discussion threads for any related problems to obtain further insight on what students believe their own conceptual deficiencies are, and to see if there were any other common issues that we may have missed within our own analysis.

Once we generated a master list of fundamental concepts for the course, we generated propositions that we were interested in assessing based on the plans that Dr. Lovric had for the course and the domain analysis of differential geometry. We constructed inventory questions that assess the relationships between concepts. I created questions independently and passed them through a preliminary content validity check; in this phase, I checked whether notation [N], terminology [T] and wording were confusing [C]. If so, the item was modified.

[1] Distractor items are options within a multiple-choice question that are incorrect, or only partially correct, but intend to address misconceptions that students may have in the material.

Proposition Assessed	Option	Inference	Content Validity			Notes
The dot product helps to measure the angle between vectors	A	Positive dot product implies the same direction vector	N	T	C	In "E"; say "linearly dependent" and not "not linearly independent" Keep; this is relevant to visualizing computations used repeatedly throughout the course.
	B	Positive dot product implies positive vector components	N	T	C	
	C	Correct; Positive dot product implies acute contained angle	N	T	C	
	D	Positive dot product implies obtuse contained angle	N	T	C	
	E	Positive dot product implies they are scalar multiples of each other.	N	T	C	

Figure 4 Sample question Analysis. Under "Content Validity", we considered Notation [N], Terminology [T], and Confusing wording [C] validity in accordance with the criticisms of the CCI presented in Gleason, et. Al, 2015.

After items were modified, they were given to the course instructor, Dr. Lovric, for review. As the instructor, he had the authority to deem whether concepts being tested were relevant and foundational, based on his plans for the course. After review, items were further revised or eliminated; on occasion, items were added to supplement conceptual relationships that may have been missed. In total, we created 91 items, including 53 items for the prerequisite concept inventory and 38 items for the in-term inventories. Of these items, we narrowed the scope of our investigation to 24 prerequisite inventory items, and 24 items for in-term inventories. These items were passed into subsequent validity tests, including cognitive labs and peer review.

We decided that all items would be multiple choice for logistical reasons; since we were planning to automate the generation of a concept map, we needed a quantifiable measure of whether a student would have successfully made the conceptual connection being measured by the question. We also wanted to ensure that the probability of a random guess being successful was equal on each question. To ensure this, we used the same number of options for each question.

Like in the development of the CCI, we elected to perform cognitive labs to validate to the inventory and theoretical framework of our investigation. We recruited McMaster student volunteers to participate in 60-minute cognitive labs based on the draft of questions developed. Students were considered eligible to participate if they were enrolled at McMaster as an undergraduate student at the time of the interview and had taken the prerequisite course sequence for MATH 3B03 as outlined by the 2019-2020 undergraduate course calendar. Of the seven students recruited, four completed the lab and elected to give us access to their data for research. Students were given access to the letter of information and consent information prior to their cognitive lab and reminded of their rights as a participant to withdraw from the interview at any time at the beginning of the meeting.

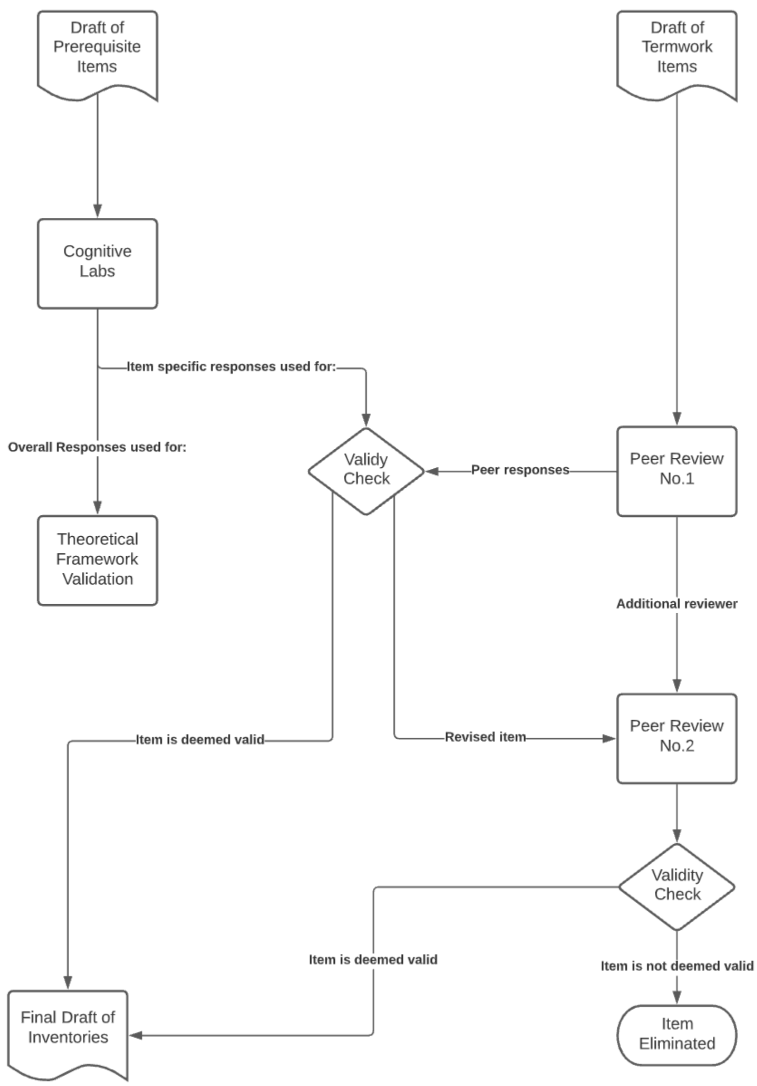


Figure 5 Overview of validity testing in Phase I.

The labs were highly structured subject-interviews, conducted virtually via Zoom or an equivalent video conferencing platform at the subjects' request. I would share my screen with a copy of the test items that were created and ask participants to "think out-loud" while solving the problem and indicate why they were selecting or not selecting each option. The role of the interviewer throughout this process was to navigate between questions, and occasionally to probe the participant for more information when their thought process was unclear. Since the participants' thought process throughout the interview was being recorded, it was important not to intervene with clarification or hints throughout the interview process unless asked for minor clarification questions. For instance, participants would occasionally ask if i represented the standard basis vector in $ai + bj + ck$, I would confirm that it was a standard basis vector. In contrast, a participant asked if i was a unit vector, I did not answer since that involved a conceptual link between the representation of the basis vector i and the length of the vector equalling 1.

All labs were audio-recorded, and then transcribed so that we could assess the problem-solving strategies and conceptual frameworks of the participants in comparison to their results on the concept inventory. The goal of this phase of research was to evaluate the quality of the questions in terms of their ability to measure the conceptual relationships that they aimed to assess, in terms of how well the item evaluates the conceptual relationship it is intended to test. For instance, if participants were misinterpreting any terms or notation used within a question but were describing a valid thought process, it was a clear indication that we needed to revise the item. Conversely, if participants were correctly identifying the solution of the question but arrived at the correct answer using invalid conceptual links, the item was revised. For some sample analysis from cognitive labs, see Appendix A.

Throughout this process, participants were allowed to take any notes they pleased, although we did not receive ethics clearance to obtain copies of their notes. This was mostly for confidentiality reasons; should students enroll in MATH 3B03, we did not want to have their handwriting on file, as we planned to collect copies of handwritten tests. However, should they be drawing or writing information, they described it to me, and I took down copies in my own writing to show the student. In addition to these occasional diagrams, I took notes manually throughout the interview, and manually transcribed the interview to verify inventory responses and student quotes within 48 hours of the completed interview. These processes were followed to ensure that the data was collected accurately, and with respect for the participants right to remain unidentifiable within the small data set collected.

Items from this phase were either passed into the final concept inventory with minor revisions or passed through another phase of peer review after major revisions along with any questions that were not validated via cognitive labs. Peer review was completed by four graduate student volunteers. Reviewers were given a copy of the item, and the desired conceptual links that were being assessed by each option within the question. They were asked then to evaluate whether the selection of each given option would be reasonable, unreasonable or neutral evidence of the described conceptual connection. If they answered, "unreasonable evidence" or "neutral evidence", they were asked to explain why they made that selection. They were also asked if they had any other comments on the item. For some sample analysis from peer review, see Appendix B.

If a prerequisite item passed through cognitive labs with minor revisions or passed through the levels of peer review after revision, it was included in the prerequisite inventory. Otherwise, the item would be eliminated, and the concept map would be modified to reflect the change in concepts we would measure. If a term-inventory item passed through two rounds of peer review without major revisions, it was included in the term work inventories. Otherwise, the items were eliminated.

Phase II: Implementation

Once items were finalized, we began recruiting student volunteers from the Fall 2020 MATH 3B03 class to participate in the research study. The inventories created in the first phase were administered as assignments in the course and marked for completion. Students were not required to give us access to their inventory data to obtain credit for completing the inventory in the course.

During the second week of synchronous class meetings, I informed the students of the study and discussed the letter of intent. Additionally, Dr. Lovric emailed the information to the entire class, posted it on the course webpage, as well as on the learning management system, thus making sure to reach all students who did not attend the synchronous meeting. A copy of the letter of information and consent questions, given, and accessible, to all students, made it clear that their academic data would be aggregated and encoded with a unique participant ID to ensure anonymity, and that they could withdraw their data from the study at any time up to December 30th, 2020. Students who chose to have their tests shared for research purposes completed a consent form, that was forwarded to course Teaching Assistants (TAs). Otherwise, students indicated whether they consented to their inventory responses to be used for research by responding to a consent question at the beginning of each concept inventory.

In addition to the concept inventories and test data, we also asked students to provide their partial birth date as indicated by their birth year, their self-reported gender identity, ethnicity and citizenship. We collected these pieces of demographic data to check if our sample was reflective of the McMaster University undergraduate student body, as outlined by the demographic overview in the McMaster University 2020 Fact Book (INSTITUTIONAL RESEARCH AND

ANALYSIS, MCMASTER UNIVERSITY , 2020) and our expected age distribution of Level III students.

The first inventory was distributed within the first week of classes, and the remaining two inventories were distributed one week prior to the first two course midterms. The course had three midterms. The data from each inventory was collected by the course TAs, who saved non-identifiable versions of the students' responses onto a secure drive. I then manually transferred each response into an excel sheet. Data entry was spot-checked at the end of the study to identify possible inconsistencies in data entry using a randomized sample of 10% of the submitted PDFs.

Modelling Cognitive Structure:

Once questions were constructed and validated in Phase I, and the instrument data was collected in Phase II, we constructed a concept map to represent the understanding of the class content as a whole. We focused on constructing a concept map for this inventory because the prerequisite concept inventory we developed measures concepts that have previously been learned by students rather than in-course knowledge. Further, the curricular implications of the concepts assessed in this inventory are of interest to us, given that one of our research questions is to assess how prerequisite courses prepare students for studies in differential geometry.

In terms of generating a concept map based on the concept inventory, we needed to find a way to represent the conceptual links that each question aims to assess. To do so, we split concept inventory questions into two categories based on the conceptual links that were being assessed: *Simple Links* and *Complex Links*.

A question is said to assess a *simple link* if there is exactly one correct option among the multiple-choice options that represents a single conceptual proposition. These questions are aiming to measure exactly one proposition and would appear as exactly one set of nodes appearing on the concept map. For instance, the item in *Example 1* would be considered a simple link style question because there is only one option that is correct, and students are not given the option to choose several options from the list provided.

Example 1: Simple Link

Suppose that $f(x)$ is continuous over $[a, b]$. Which of the following is always true?

a. $\int_a^b f(x)\,dx$ is the area of the region bounded by the graph of $f(x)$, the x-axis, and the lines $x = a$ and $x = b$.
b. $\int_a^b f(x)\,dx$ is finite.
c. $\int_a^b f(x)\,dx$ is the antiderivative of $f(x)$
d. $\int_a^b f(x)\,dx$ may not exist.
e. None of these.

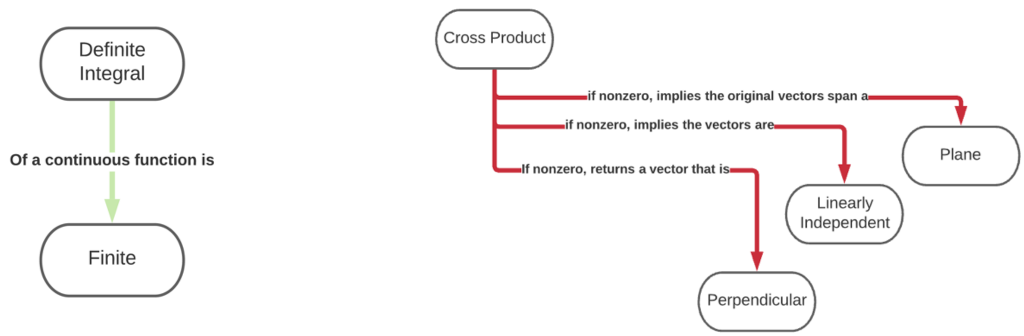
At times, we elected to test multiple propositions at the same time by using a single item test related concepts within its set of options. A question that assesses multiple concepts at the same time is a *complex link*. These items have one fully correct option within the set that represent a conceptual link between two or more concepts, and other options that signify a subset of the conceptual links made by the fully correct option. By choosing to include complex links, we were able to cover more concepts that were deemed relevant to the course without adding too many questions the inventory. We did so in an attempt to limit survey fatigue from student participants, and to allow for partially correct answers to be represented on the concept map. We felt that this was an appropriate choice since conceptual understanding is not necessarily a binary variable and limiting the number of items would encourage student participants to finish their inventories. An example of such an item is seen below.

Example 2: Complex Link

Suppose that we have two vectors, $u, v \in \mathbb{R}^3$ such that their cross product $v \times u$ is a nonzero vector w. Then:

a. w is perpendicular to both u and v
b. u and v span a plane
c. u, w and v are linearly independent vectors.
d. (a) and (b)
e. (a), (b) and (c)

This item's correct answer is (e), which includes three simple links between concepts; specifically, the propositions being assessed are that a nonzero cross product implies that the resultant vector is perpendicular to the original two vectors, that a nonzero cross product implies that the original two vectors span a plane, and that a nonzero cross product implies that the original two vectors are linearly independent. Should a student select (e), this is considered evidence that each of the previously mentioned conceptual links have been made, since they chose (e) from a list of conceptually distinct and viable distractor items, and all three links would appear on the concept map. On the other hand, if a student chose (a) in this item, we took this to mean that the student had presented evidence of the first conceptual link mentioned, despite not providing evidence of the other two propositions; this means that while they would receive a score of zero on the concept inventory assignment question, we would still consider the students' partially correct structure in the concept map.



(i)
Simple link resulting from a selection of (b) in example (1).

(ii)
Complex link resulting from a selection of (e) in example (2).

Figure 6: Translation from sample items to their representation on a concept map

We constructed the expert-level concept map from our finalized concept inventory items and used LucidChart to generate the concept map (Lucidchart, 2020). In the map, we denote knowledge that is developed in the calculus course sequence in green arrows (as in Figure 6(i)), and knowledge that is developed in the linear algebra sequence in red arrows (as in Figure 6(ii)). We chose LucidChart because there is an option to upload a CSV file to auto-generate the map based on the shape of the data. We referred to the uploaded structure, including each node and link from each question and its' associated assumptions, as the *base map*, seen later in Figure 13. Once we had the base map established and had decided on a scoring system for each proposition, we could automate the data linking to each node on the concept map.

Using this method, we constructed three additional concept maps based on the data we received from student participants. Each of these maps corresponds to different success criteria required to construct a link between concepts. We defined the success criteria in terms of the number of students who participated to correctly identify a conceptual link. For instance, in our "Quartile 1" map, we allowed a conceptual link to be made if at least 25% of students who participated provided evidence that they had the cognitive structure being assessed. For a simple link item, this would mean that at least 25% of students would have to identify the only correct option for the item, whereas for a complex link item, this would mean that at least 25% of students would have had to identify the options that were considered evidence of that given conceptual link. For instance, in Example 2, we added the conceptual link assessed in item (a) if the sum of students who answered (a), (d) or (e) all added to at least 25% of participants, since each of these items provides evidence that the student has the cognitive structure associated with option (a). We repeated this process to construct a "Quartile 2" map, where at least 50% of respondents were required to provide evidence of the conceptual link for it to appear on the map, and a "Quartile 3" map, where at least 75% of respondents were required to provide evidence of the conceptual link for it to appear on the map.

Inventory Analysis:

To gain further insight on what students are doing well in, and what areas could be improved, we turned to item and distractor analysis. This is as a method to analyze students' performance on the concept inventory as a whole, and on individual items. In addition to offering valuable information on students' conceptual understanding of the material, this technique allowed us to statistically validate test items, and to test for internal validity. We used the difficulty and discrimination index to give us an indication of which questions students found most challenging and which questions were able to discriminate between high and low achieving students. The difficulty index p is the ratio of the students who were successful on the item compared to the number of students who responded to the item. We followed the following table for classifying items as "difficult" or "easy".

Item Difficulty Index p	**Interpretation**
$0.6 \leq p$	Easy Item
$0.4 \leq p < 0.6$	Moderate difficulty
$p < 0.4$	Difficult Item

Figure 7: Framework for interpreting the difficulty index p, *adapted from (Bai & Ola, 2017).*

The discrimination index D for a particular item is given by the equation below, where $\overline{X_C}$ represents the mean score for students who answered an item correctly, $\overline{X_W}$ is the mean score for students who answered the item incorrectly, σ is the standard deviation of the total inventory score, and p is the item difficulty index:

$$D = \frac{\overline{X_C} - \overline{X_W}}{\sigma}\sqrt{p(1-p)}$$

This score is meant to describe how accurate an item is in distinguishing between students who score highly on an assessment and those who do not. Positive values of D imply that the students who performed well on the item also performed well on the inventory as a whole, indicating that stronger students were comparatively more successful on the item; conversely, negative values of D indicate that comparatively weaker students performed better on the item than strong students, indicating that the item does not effectively identify strong performing students (Bai & Ola, 2017). Items are considered to be highly discriminating if their discrimination index is greater than 0.4, and ineffectively discriminating if their index is less than 0.3.

Discrimination Index D	**Interpretation**
$D \geq 0.4$	Item effectively distinguishes between high and low scoring students
$0.3 \leq D < 0.4$	Reasonably good at discriminating between high and low scoring students.
$D < 0.3$	Item ineffectively discriminates between high and low scoring students.

Figure 8: Framework for interpreting the discrimination index, adapted from (Bai & Ola, 2017)

For our purposes, we have adapted the interpretation frameworks outlined in Bai and Ola's paper based on their recommendation that parameters for p and D are based on the purpose of the test itself (Bai & Ola, 2017). Since our goal is to assess cognitive structures, we will consider an item with a difficulty index between 0.4 and 0.6 ideal because it will exclude questions that are too difficult or too easy to reasonably assess this variable. We also are interested in highly discriminating questions, since this will offer insight on what concepts high achieving students have mastered compared to low achieving students.

With these two measurements in mind, we created the following table for interpreting the item scores and split items into categories, depending on how difficult and discriminating an item is measured to be.

		Discrimination Index D		
		$0.4 \leq D$	$0.3 \leq D < 0.4$	$D < 0.3$
	$p < 0.4$	High difficulty and highly discriminating power. (1)	High difficulty and moderate discriminating power. (2)	High difficulty and low discriminating power. (3)
Difficulty Index p	$0.4 \leq p < 0.6$	Ideal difficulty for cognitive structure and high discrimination power. (4)	Ideal difficulty and moderate discrimination power. (5)	Ideal difficulty with low discrimination power. (6)
	$0.6 \leq p$	Low difficulty and high discrimination index (7)	Low difficulty and moderate discrimination power (8)	Low difficulty and low discriminating power (9)

Figure 9: Interpretive framework for difficulty and discrimination indices used in analyzing item responses from inventories.

Items in categories (1), (3), and (9) are of particular interest to our investigation, as they represent extreme cases. For instance, if an item is categorized in item (1), it would mean that not only was the item highly correlated with high scores on the inventory, but it was also answered correctly by relatively few students; from this, we can conclude that this cognitive structure is only possessed by top performing students. In contrast, an item categorized in (9) would indicate a base cognitive structure that is possessed by both strong and weak performing students.

Items in category (3) are those items that were neither answered correctly by at least 40% of students, nor effectively discriminate between high and low scoring students. If the item can be shown as valid through the cognitive labs and/or peer review, then this indicates a cognitive structure that is absent from strong and weak performing students alike, thereby representing a commonly misunderstood concept for most students.

Lastly, to check for internal consistency, we use the $KR20$ reliability coefficient, which assesses how consistent scores on items within a test are with each other. We use $KR20$ as opposed to $KR21$, as it does not assume that all items are of equal difficulty (Frankel & Wallen, 2009); as we have a mix of single link and complex link style questions, this measure was most appropriate in determining internal consistency. To calculate $KR20$, we use the following formula:

$$KR20 = \frac{k}{k-1} \times \left(1 - \sum_{i=1}^{k} \frac{p_i(1-p_i)}{\sigma^2}\right)$$

Where k is the number of items on an assessment, p_i is the difficulty index of the i^{th} item, and σ is the standard deviation of the total test scores (Bai & Ola, 2017). A $KR20$ score greater than 0.5 is considered acceptable in teaching practice, as it indicates that items within an assessment are related to each other (Bai & Ola, 2017); however, the standard for research in education is a $KR20$ score of at least 0.7 (Frankel & Wallen, 2009).

Results: Phase I

In this phase of research, we created and collected evidence of validity for a prerequisite inventory of twenty-two conceptual items and four demographic questions, and two concept inventories based on the in-term material that had nine questions each. In addition to providing evidence of item validity, the cognitive labs from this phase provided evidence of validity for theoretical framework for the study as a whole. This phase of research also identifies what concepts are necessary to assess for our sample, articulate the assumptions on student knowledge were required in our model of cognitive structures, and to generate the concept map associated with the pre-requisite concept inventory.

Concept Inventory Development:

When constructing the pre-requisite concept inventory, we were surprised at how the majority of concepts required to build a rich conceptual framework relied on the first-year level content. The development of concepts within the course-texts were explicit in drawing connections between previously learned mathematics concepts and objects to newer ones; as such, much of the content we included in the concept inventory included basic calculus results. If not explicitly used in exercises or proofs, we often saw that the fundamental calculus concepts were used in developing intuition and reasoning about objects introduced in the course. For instance, consider the passage from Pressley's *Elementary Differential Geometry*, below that introduces and defines arclength, where $\mathcal{C}$ denotes the image of a curve parameterized by $\gamma(t): \mathbb{R} \to \mathbb{R}^2$:

[…]

If we want to calculate the length of a (not necessarily small) part of $\mathcal{C}$, we can divide it into segments, each of which corresponds to a small increment of δt in t, calculate the length of each segment using (1.4), and add up the results. Letting δt tend to zero should give the exact length.

This motivates the following definition:

Definition 1.2.1: *The* arc-length *of a curve γ starting at the point $\gamma(t_0)$ is the function $s(t)$ given by:*

$$s(t) = \int_{t_0}^{t} ||\dot{\gamma}(u)||du$$

(Pressley, 2012)

From this passage, we can see that the new object "arclength" is motivated based on students' conceptual understanding of tangent vectors, Riemann Sums, and the definition of the integral. Without a foundational understanding of each of these objects, the intuitive understanding of arclength as the "sum of lengths" would be inaccessible to novice students. As such, we included inventory questions that relied on the conceptual understanding of these objects.

We based our question pre-requisite concept inventory items on material emphasized in the first three chapters of the course text. A summary of concepts in the finalized inventory is shown below.

Q#	Content	Q#	Content	Q#	Content	Q#	Content
1	Properties of functions	**2**	Mean value theorem	**3**	Existence of an inverse	**4**	Properties of integration
5	Representation of the definite integral	**6**	Physical applications of definite integrals	**7**	Interpretation of Riemann Sums	**8**	Scalar multiplication of a vector
9	The length of unit vectors	**10**	Dot product, linear independence and span	**11**	Dot product and comparing directions of vectors	**12**	Nonzero cross products, span and linear independence
13	Projection of perpendicular vectors	**14**	The result of a projection of two vectors	**15**	Determinants and invertibility	**16**	Partial derivatives and slope
17	Linearization and approximation	**18**	Representation of double integrals	**19**	The derivative and approximation	**20**	Interpreting values of derivatives and the chain rule
21	Visualization of Riemann sums	**22**	Equation of a plane and the span of vectors				

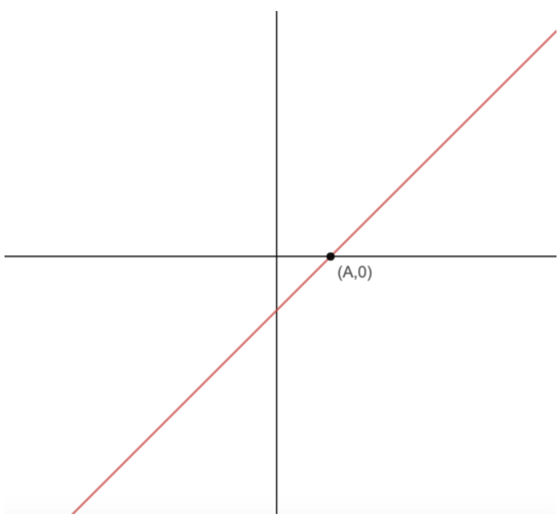
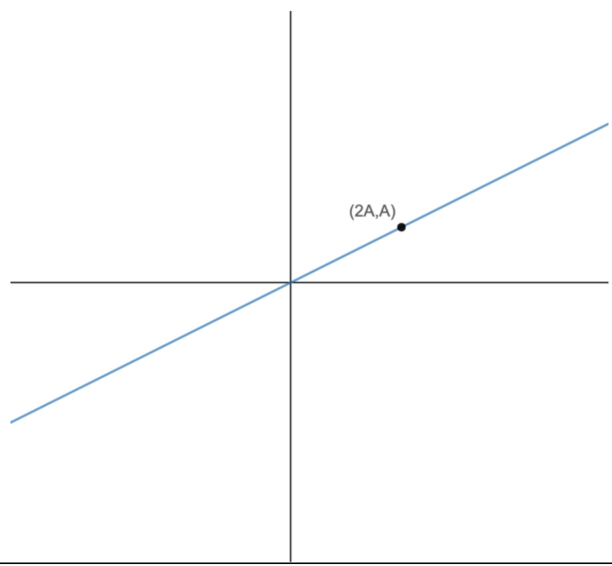
Figure 10: Item topic coverage for prerequisite concept inventory.

Interestingly, we rarely saw computational elements from calculus in the text as necessary to support the learning of new material. Though necessary for success in exercises, the computational understanding required for learning the concepts of the course were primarily based in properties of the objects themselves. For instance, the text required students to understand when vectors were linearly dependent based on the result of their cross product in order to develop propositions, but the exact cross product was not computed in these cases (however, it was computed, using its properties, in other cases). Therefore, we were most interested in developing questions that rely on conceptual knowledge and understanding and are independent of skill.

In other words, we were interested in items that would be able to discriminate between students who simply knew what objects were, and those who were capable of reasoning with these objects in a spatial-visual or analytic context. For instance, consider the item below that assesses the value of a derivative, the chain rule and interpretation of critical points[2]. Computational proficiency with derivative rules is necessary to successfully complete this question, as is an understanding of the sign of products of integers; however, the ability to compute these elements without a conceptual understanding of derivatives and interpret the sign of the derivative from the spatial context of the item is unlikely to be successful on the item. Therefore, this item tests how students work with and interprets derivatives, composition and derivative rules, rather than exclusively testing computational knowledge.

Example 3:

Suppose that $f: \mathbb{R} \rightarrow \mathbb{R}$ and $g: \mathbb{R} \rightarrow \mathbb{R}$ are differentiable functions, and the graphs of their derivatives are shown below:

$f'(x)$	$g'(x)$
	

If $f(x) > 0$ and $g(x) > 0$ for all $x \in \mathbb{R}$, then which of the following are true?

a. The function $g(x)$ is increasing for all values of x
b. The function $f(x)$ has a critical point at $x = A$
c. The derivative of $g(f(x))$ is negative for $0 < x < A$
d. (b) and (c) only
e. All of the above

[2]The original item and associated transcripts from cognitive labs are in the Appendix A, for reference.

Cognitive Labs:

Of the 53 questions developed for our prerequisite concept inventory, we performed cognitive labs on twenty-four items. We recruited volunteer undergraduate students in the mathematics department to participate in these labs. We received seven inquiries from participants, of which four agreed to complete the cognitive lab in the study and have their transcripts kept for research.

We used participants' transcripts of their work to assess whether the question being posed to participants was effectively measuring the conceptual framework it was intended to measure. This could be asserted in one of two ways; participants who answer an item correctly describe a valid problem-solving strategy and make correct conceptual connections, or participants who answer an item incorrectly follow an invalid problem-solving strategy or make incorrect conceptual connections. In either of these cases, the item would have successfully identified participants as having the appropriate cognitive structures associated with the item, and we could validate the item[3].

Each test item was evaluated individually. From these labs, we were able to gather enough evidence of validity for fifteen items without need of further peer review or revision. Five items were considered valid after minor revisions, like notational changes or rephrasing options such as "All three" to "(a), (b) and (c)" for consistency in formatting. The remaining four items were passed into peer review after major revisions.[4]

Once we examined each item on their own, we considered the test as a whole to check the instrument we constructed was in line with the theoretical framework for our investigation. The majority of questions focused on testing the understanding of connections between concepts, rather than testing computational proficiency. However, participants frequently attempted to directly compute an answer within the item, even when such a direct computation was irrelevant or unnecessary.

We observed that participants that had strong existing structures to shift to problem solving performed better than those who had to rely on pre-existing structures. When participants used the words "remember" or "recall", they were generally less successful on an item than those who did not need to rely on memory. Participants who engaged in either of these behaviours

[3] For sample item validation tests, see Appendix A.
[4] For sample peer feedback and methodology, see Appendix B.

were often unable to develop an alternative problem-solving process when they had difficulty answering questions.

Some references to the past should be expected, since we were testing concepts that were already acquired by participants; however, when participants were able to reason about concepts *without* situating them in the past, this indicated that they still had a functional cognitive structure to work with for their problem solving, and thus they were more successful on items presented to them. The cognitive lab participants CL3 and CL2 very consistently referenced previous courses, "remembering" aspects of the questions presented to them, positioning their knowledge in the past.

We also found that even when participants had taken more courses in mathematics, they were not necessarily more successful on the inventory. When they mentioned later coursework and attempted to use their it to justify their reasoning, they were often unsuccessful in their attempts to solve the inventory problems. Among the participants who have taken further mathematics courses beyond the Level I and Level II prerequisite courses for MATH 3B03, the references to these courses in the past was also an indicator for limited success on the inventory question, even though the material in those courses was, in some ways, more advanced.

For instance, CL1 attempted to use techniques from Real Analysis when trying to interpret a Riemann Sum, referring to the infimum and supremum when attempting to interpret a Riemann Sum, and the Frenet Frame when faced with a problem on the cross product; though this participant had taken the course before, they were unable to recognize the relevant concepts and relationships necessary to analyze the item they were given. This is consistent with Ausubel's Theory of Assimilation; when students are unable to attach new concepts to an appropriate conceptual anchor, the new concept is vulnerable to being forgotten, or is remembered in a way that doesn't allow for rich problem solving.

The two themes of time-situated knowledge and computational problem-solving appeared to be indicators of success with the instrument as a whole. After transcribing the recordings, we counted the number of references to the past or explicit attempts to compute when there was no need to compute in the item in Figure 11.

Participant	Time-Situated References[5]	Computational Problem-Solving References[6]	Items Seen	Items Attempted	Score
CL1	18	8	14	13	6
CL2	4	4	24	24	20
CL3	18	9	24	22	12
CL4	46	15	24	18	10

Figure 11 Summary of references to the past and computation with respect to final score and items attempted

As an aside, we also saw participants reference exams and courses that they did well in academically, but then incorrectly identify the option in a question or follow an incorrect train of thought while solving a problem with that course content. Participants who mentioned this and were incorrect did so with verbal evidence; they did not have the cognitive structure necessary to successfully respond to questions when given to them but referenced high marks on their exams for the relevant courses. It's worth noting that this is consistent with the findings from the Institute for Machine and Human Cognition (Institute for Machine and Human Cognition, 2003); that is, while conceptual understanding and academic performance are related variables, the presence of a strong academic performance does not necessarily indicate the presence of a strong cognitive structure of the courses' concept domain.

Though the sample size for the cognitive labs was admittedly small, the purpose of the lab was to evaluate the validity of the questions at measuring the conceptual frameworks intended, not to draw any conclusions about the cognitive structures of Level III students. To obtain this, we needed a high level of detail in student thought processes while responding to the items, rather than a large sample with test-scores; therefore, we can say that the results of the cognitive labs are valid, and that the validation and remediation of test items is justified with the data obtained from the labs.

[5] We used the following search terms when counting each indication of time-situated knowledge: "recall", "remember", "studied", "just did this", "did these", "exam". If two relevant search terms were in the same sentence, they were counted once, as was any reference to a previous instructor or course code.

[6] We used the following search terms when counting each indication of computational problem solving: "formula", "equation", "compute", "solve". We also counted whenever a participant engaged in literal computation, and subtracted references to computation where it was an appropriate step in the overarching problem-solving structure of the question.

Concept Map and Assumptions

Once the prerequisite concept inventory was finalized, we established a base map that models all conceptual connections tested for in the inventory, seen in Figure 13. When establishing what the inventory was assessing, we realized that we had to impose some assumptions implicitly within the phrasing of the questions. For instance, in the item below, the phrasing of the question assumes that a double integral is computed over some domain of integration. In the phrasing of each of the options, we include information of the domain $\mathcal{D}$ for which the function is being integrated over; as a result, we cannot conclude from this question whether or not the student understands that double integrals are given over some two-dimensional domain.

Example 4:

Suppose that $f(x,y): \mathbb{R}^2 \to \mathbb{R}$ is a nonnegative and continuous function on $[0,1] \times [0,2]$, and consider the following double integral:

$$\int_0^1 \int_0^{2x} f(x,y)\, dy\, dx$$

Then:

a) The integral represents the volume of the solid bounded by the surface $f(x,y)$ and the plane $z = 0$ over some domain $\mathcal{D}$
b) The integral represents the surface area of the surface $f(x,y)$ over some domain $\mathcal{D}$
c) The domain of integration $\mathcal{D}$ is a triangle with vertices $(0,0), (1,0), (1,2)$
d) (a) and (c)
e) (b) and (c)

Instead, we must take it as an assumed knowledge, based on how the question is phrased. This assumption is indicated on the model by coloring the connection between these concepts with blue; an example for this item translated into a structure is shown below.

Assumption	**Graphical link**
The double integral is computed over a domain of integration in the xy plane.	

Figure 12 Sample Assumption Notation

Lastly, we classified knowledge that originates from calculus and linear algebra prerequisite sequences and denoted it on the concept map using green for the calculus sequence, and red for the linear algebra sequence. The base map for this structure is shown on the following page. Due to the validation of the instrument items and the extensive domain analysis on the items we created, this model can be seen as a reflection of the key aspects of prerequisite courses. Though not exhaustive, the connections modeled between concepts are indeed relevant and necessary for the course and are effectively being measured by the test items.

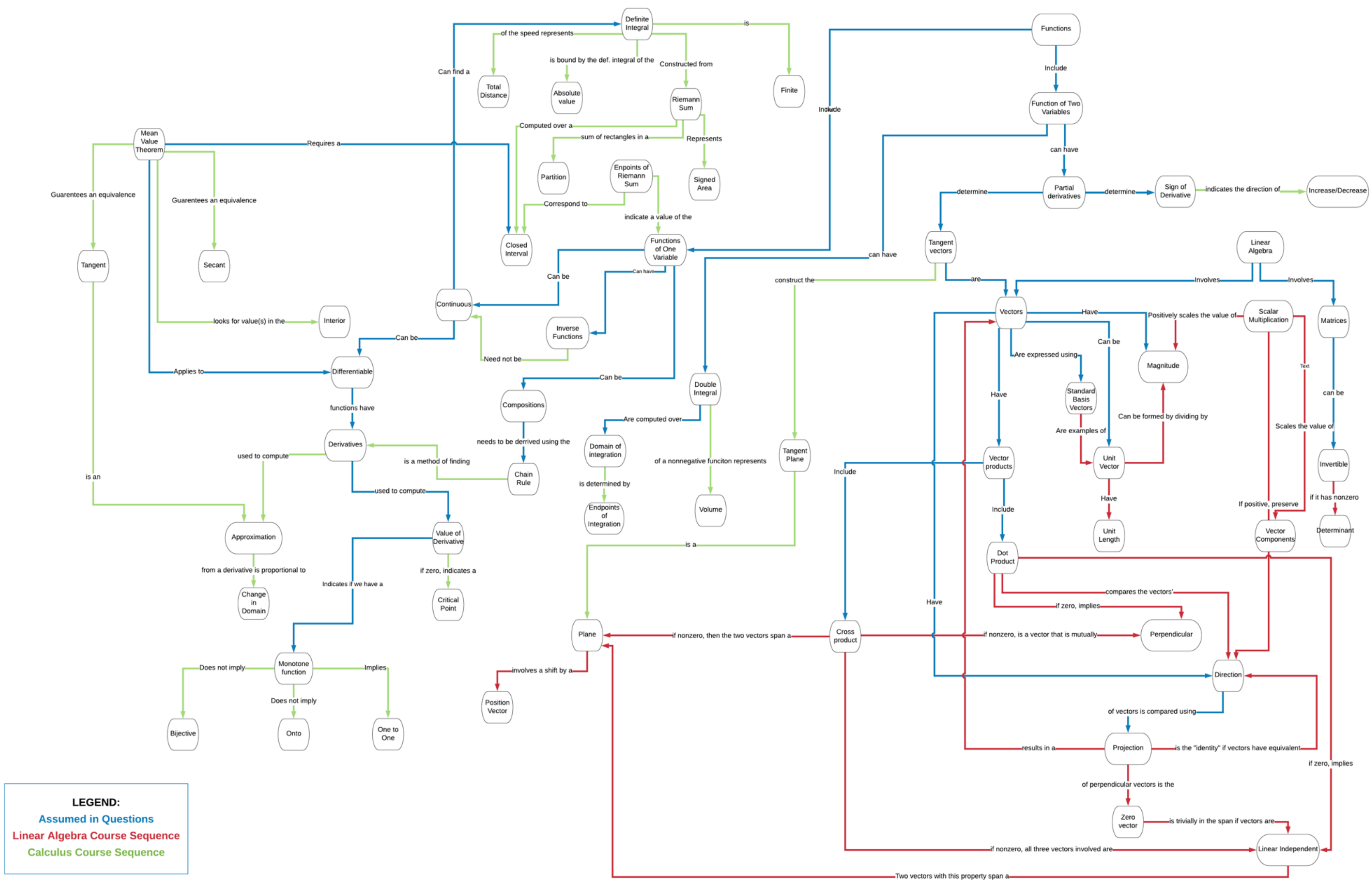


Figure 13: Base Concept Map for Pre-Requisite Concept Inventory. Available online [here](https://lucid.app/lucidchart/15032623-43cc-4084-b923-a4202344e406/view), using the URL: https://lucid.app/lucidchart/15032623-43cc-4084-b923-a4202344e406/view

Results from Phase II:

Of the 139 students enrolled in MATH 3B03, we obtained data from 129 participants over the course of the study. This section contains the results from each of the three concept inventories, and test information for the participants involved.

Sample:

In the first inventory, we obtained demographic information on the sample from 118 participants. Participants were not required to answer any of the demographic questions, and they were not factored into the inventory score or participation marks for the inventory. Participants were also given the opportunity to disclose their demographic data in free-response format, rather than through a given selection of options, to ensure that they were able to express the identity that most aligned with them. We saw a relatively even distribution of male and female identifying students within the sample, with a 57:51 ratio of male to female students. No other gender identities were reported.

	Male	Female	Respondent Total
Count	57	51	108
Percentage of Respondents	53.8%	47.2%	

Figure 14 Gender distribution, $n = 108$.

The percentage of female students in the sample is lower than the percentage of female students at McMaster at a whole, with 54.8% of undergraduate students at McMaster reporting a female gender identity in the 2019-2020 academic year (INSTITUTIONAL RESEARCH AND ANALYSIS, MCMASTER UNIVERSITY , 2020).

We also saw that the majority of the students in the sample were international students, with 36.6% reporting Canadian citizenship. This disparity between domestic and international students is especially startling when considering that the proportion of undergraduate international students for McMaster University as a whole for the 2019-2020 academic year was 13.2% (INSTITUTIONAL RESEARCH AND ANALYSIS, MCMASTER UNIVERSITY , 2020). However, this average is not representative of individual faculties, as interests of international students are not evenly distributed across all disciplines.

	Canadian	Dual Citizenship	International	Respondent Total
Count	37	4	60	101
Percentage of Respondents	36.6	4.0	59.4	

Figure 15 Citizenship distribution of the sample, $n = 101$

The age distribution, on the other hand, is well in line with what we expected for the sample of Level III students. Since the course is available to students in Level III and Level IV, we expected the average age of participants to be twenty-one, assuming that students enroll in post-secondary after completing secondary school at eighteen. We used a partial birth date to estimate the ages of sample participants and found that most of the sample did have partial birth dates indicating an age between twenty and twenty-two.

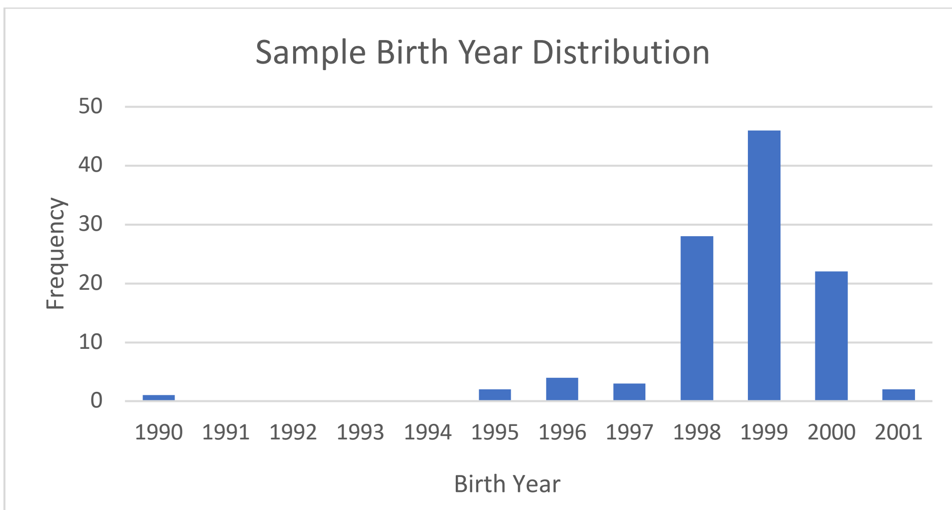


Figure 16 Age distribution as indicated by partial birthdate, $n = 108$*.*

Inventory 1:

The first concept inventory had an average score of 9.08 out of 22, and a standard deviation of 4.69. Though this appears low, it's worth noting that we expected the majority of items to have a difficulty index between 0.4 and 0.6 due to the number of conceptual links being tested within a single test item, so the mean score of 9.08 is along the lines of what we would hope for with this inventory. A summary of each question and its associated p and D values are shown in Figure 17.

Question #	Discrimination Index	Difficulty Index	Question #	Discrimination Index	Difficulty Index
1	0.275	0.382	12	0.453	0.400
2	0.403	0.418	13	0.486	0.455
3	0.194	0.455	14	0.371	0.245
4	0.474	0.482	15	0.367	0.345
5	0.148	0.118	16	0.396	0.527
6	0.382	0.591	17	0.279	0.273
7	0.432	0.309	18	0.320	0.418
8	0.439	0.427	19	0.317	0.509
9	0.456	0.573	20	0.416	0.355
10	0.350	0.582	21	0.451	0.464
11	0.502	0.545	22	0.343	0.236

Figure 17 Summary of item analysis from Inventory 1, based on 110 valid responses.

On average, the difficulty index and discrimination index for items on this inventory was 0.41. We also found that there were three items in category (1) and three items in category (3), with the majority of the items found to be highly discriminating, and with difficulty index between 0.4 and 0.6. For a summary of items in each category, see Figure 18.

Difficulty Index p		Discrimination Index D: $0.4 \leq D$	Discrimination Index D: $0.3 \leq D < 0.4$	Discrimination Index D: $D < 0.3$
	$p < 0.4$	**Q7:** Interpretation of Riemann Sums **Q20:** Interpreting values of derivatives and the chain rule	**Q14:** The resultant of a projection of two vectors **Q15:** Determinants and invertibility **Q22:** Equation of a plane and span of vectors	**Q1:** Properties of functions **Q5:** Representation of the Definite Integral **Q17:** Linearization and approximation
	$0.4 \leq p < 0.6$	**Q2:** Mean Value Theorem **Q4:** Properties of Integration **Q8:** Scalar multiplication of a vector **Q9:** Length of unit vectors **Q11:** Dot product and comparing directions of vectors **Q12:** Nonzero cross products, span and linear independence **Q13:** Projection of a vector **Q21**: Visualization of Riemann Sums	**Q6:** Physical applications of definite integrals **Q10:** Dot product, linear independence and span. **Q16:** Partial derivatives and slope **Q18:** Representation of Double Integrals **Q19:** The derivative and approximation	**Q3:** Existence of an Inverse
	$0.6 \leq p$	***None***	***None***	***None***

Figure 18: Interpretation framework for items in inventory 1.

Because the inventory was designed to highlight misconceptions for students, we can use the distractor options in each item that is in these significant categories for an indication of what misconceptions are acting as barriers for lower achieving students, and what kind of conceptual issues are prevalent among all students. A summary for relevant item numbers and their distractor selections are shown in Figure 18.

Category	Question Number	A	B	C	D	E	Total Responses
Category 1: High Difficulty and Highly Discriminating	**7**	15	28	3	26	35*	107
	20	7	28	8	40*	27	110
Category 3: High Difficulty and Ineffectively Discriminating	**1**	15	42*	4	16	32	109
	5	69	14*	19	5	3	110
	17	21	18	7	31*	32	109

*Figure 19 Distractor Summary for items in categories (1) and (3). Note that "total responses" here refers to the number of responses that come from valid inventory submissions; that is, responses from participants who answered at least 19 of the 22 items. The correct response is indicated by *.*

The scores of the inventory were used to calculate the $KR20$ coefficient, which gave a value of 0.766, indicating that the instrument has internal consistency. For reference, the minimum value for instruments in educational research is 0.7. Lastly, we modeled the class level cognitive structures using quartile maps based on the percentage test score. That is, the Quartile 1 map represents the conceptual connections that *at least* 25% of participants provided evidence of, and so on, seen on pages 41 through 43, with each representing a particular quartile.[7]

[7] All concept maps include a URL that links to the web-based version of the map, which is interactive, should there be challenges with viewing and reading the propositions on the map here.

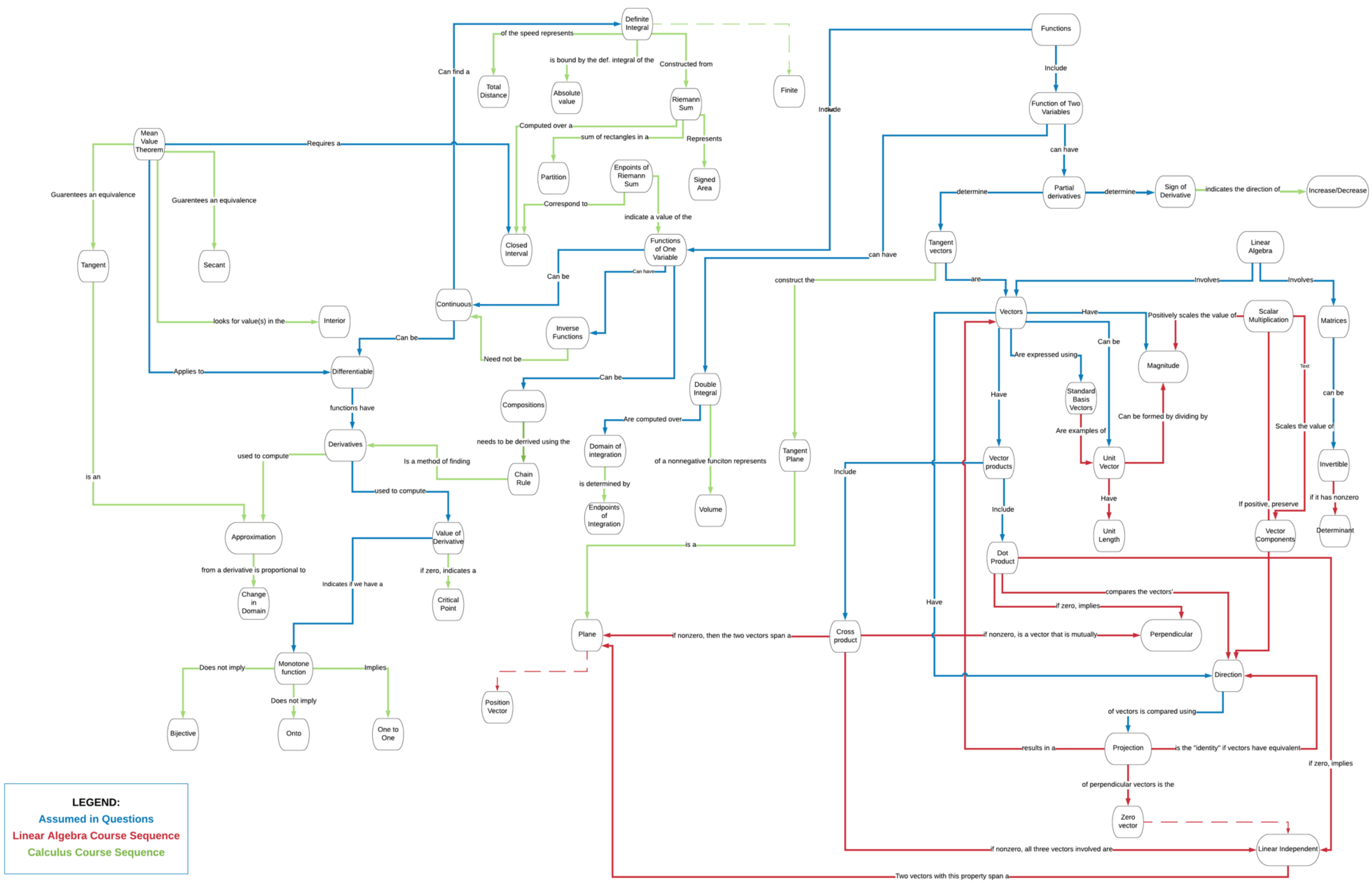


Figure 20 Quartile 1 Map for Inventory 1, using cut off of 28 correct answers. Available online [here](https://lucid.app/lucidchart/15032623-43cc-4084-b923-a4202344e406/view), using the URL [*https://lucid.app/lucidchart/15032623-43cc-4084-b923-a4202344e406/view*](https://lucid.app/lucidchart/15032623-43cc-4084-b923-a4202344e406/view)

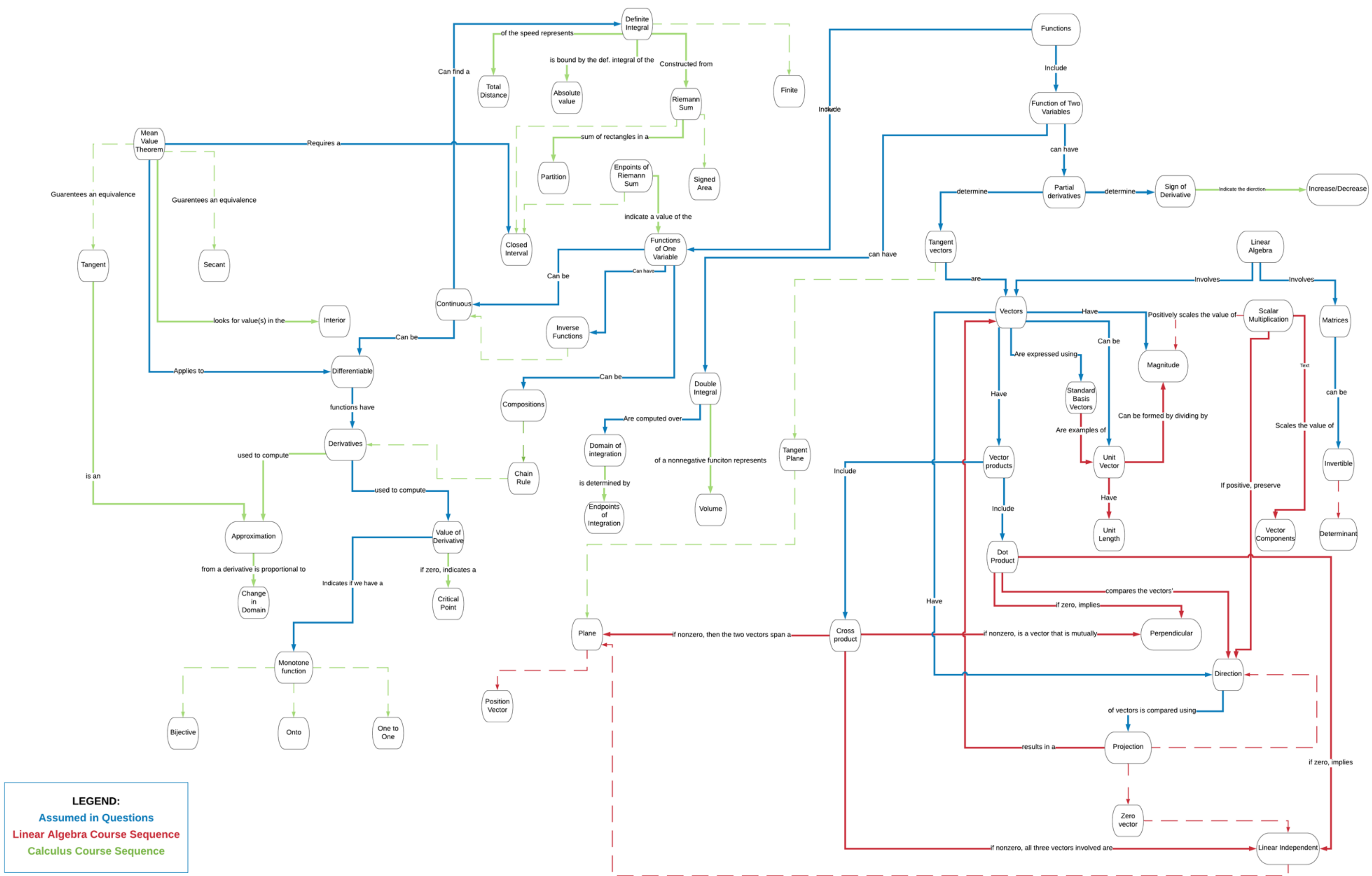


Figure 20 Quartile 2 map for inventory 1, using a cut off of 55 correct answers. Available online here, using the URL https://lucid.app/lucidchart/15032623-43cc-4084-b923-a4202344e406/view

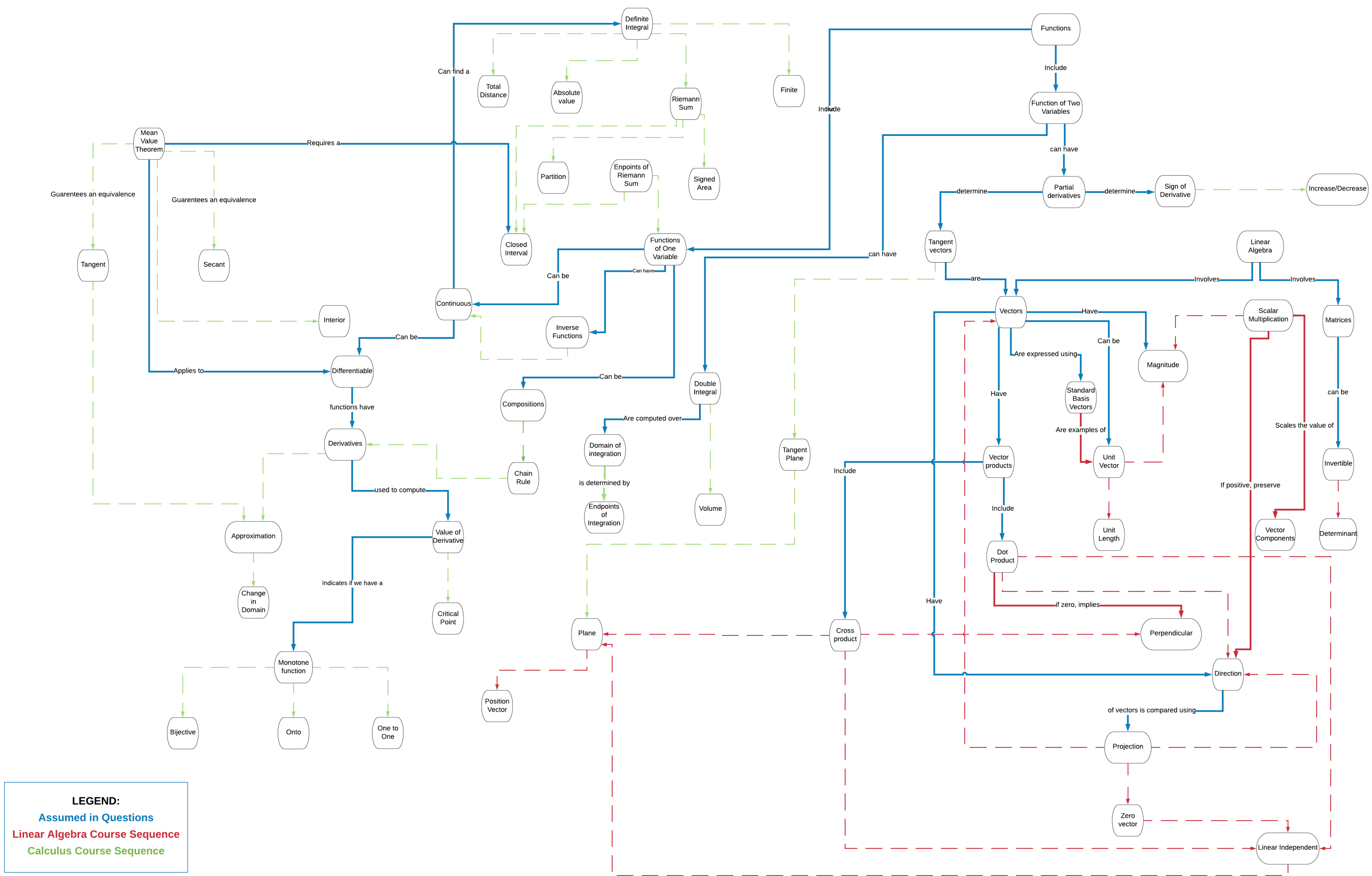


Figure 21 Quartile 3 map for inventory 1, using a cut off score of 83. Available online here, using the URL: https://lucid.app/lucidchart/15032623-43cc-4084-b923-a4202344e406/view

Inventories 2 and 3:

The second concept inventory had an average score of 3.91 out of 9, and a standard deviation of 1.65. We also saw a decrease in valid responses, from 110 in inventory 1 to 96 in inventory 2.

Question #	Discrimination Index	Difficulty Index	Question #	Discrimination Index	Difficulty Index
1	0.769	0.594	**6**	-0.115	0.167
2	0.115	0.063	**7**	0.615	0.604
3	0.615	0.375	**8**	0.038	0.125
4	0.577	0.698	**9**	0.577	0.708
5	0.769	0.573			

Figure 22 Summary of item analysis for Inventory 2, based on 96 valid responses

The third concept inventory had an average score of 4.11 out of 9, and a standard deviation of 2.04. The drop in responses continued from 96 valid responses for inventory 2, to 86 valid responses for inventory 3.

Question #	Discrimination Index	Difficulty Index	Question #	Discrimination Index	Difficulty Index
1	0.519	0.517	**6**	0.438	0.427
2	0.502	0.360	**7**	0.389	0.674
3	0.405	0.393	**8**	0.455	0.337
4	0.035	0.236	**9**	0.478	0.742
5	0.519	0.427			

Figure 23 Summary of item analysis for inventory 3, based on 89 valid responses

Neither in-term concept inventory met the standards for educational research, with a $KR20$ value of 0.217 for inventory 2, and 0.519 for inventory 3. This could be due to the smaller inventory size compared to inventory 1, or the survey fatigue that we encountered throughout the study. We also believe that the reason we obtained smaller $KR20$ values may be because the timing for distributing the instruments. Inventory 1 attempted to measure established cognitive structures, whereas these inventories attempted to measure the development of them.

Correlation to Academic Achievement

Throughout the term, we collected student midterms and recorded their final scores. This was largely to investigate whether we could see a measurable relationship between inventory achievement and academic achievement. The sample size of students who allowed us access to their midterm grades started low and decreased throughout the semester. This may be because students were embarrassed by their scores after the first midterm, which had an average of 49.43%, or that students were not incentivized to provide a copy of their midterm for credit. It may also be because we calculated test the aggregate test score grade based on the students who completed all three term tests, which limited the number of participant midterm grades that we could include.

However, we computed the correlation between inventory scores and test scores anyway to see if any correlations were compelling enough to warrant further investigation. We found that the correlation between students' first inventory score and first test was 0.57594816 ($n = 29$), and that the correlation between students' first inventory scores and their aggregate test scores was 0.93076153 ($n = 11$). The correlation between students' second inventory and test scores was 0.74906192 ($n = 13$), and the correlation between the students' third inventory and test scores was 0.28349892 ($n = 11$).

Discussion and Analysis:

We begin our discussion of results by examining the concept map generated and exploring the inventory questions that correspond to extreme discriminating and difficulty index values. These two measures afford us a general overview of the class' calculus and linear algebra achievement, and a targeted analysis of any misconceptions that are prevalent at the class level. In combination, these methods of analysis allow us to assess the curriculum's ability to prepare students for study in upper-level mathematics courses and offer insights on what course sequences may benefit from further review.

Map Analysis:

From the Quartile 1 map, we observe that less than 25% of students were unable to identify that a definite integral of a continuous function was necessarily finite, and unable to identify that the zero vector is not in the span of linearly independent vectors, and had difficulty identifying the role of a vector without a parameter coefficient in the equation of a plane.

Moving from Quartile 1 to Quartile 2, we begin to see concepts being, the majority of which come from the calculus course sequence. Ideas of one-variable tangents, approximation and derivatives are preserved; however, the conceptual deficiencies that arise when moving the success criterion from 25% to 50% are as follows:

1. The Riemann Sum in one variable represent signed area and are computed over closed intervals.
2. Monotone functions are not necessarily bijective, or onto—but they must be one to one.
3. The linearization of a function of two variables is a plane, and tangent vectors are used to construct a tangent plane.
4. Inverse functions need not be continuous.
5. Application of the Mean Value Theorem guarantees an equivalence of the slope of the secant over a closed interval to the slope of a tangent to the curve within the interval.

The majority of basic connections between concepts within linear algebra appear to be largely unaffected by the higher standard for class success, including the effect of scalar multiplication, definition of a unit vector, along with various properties and consequences of the dot and cross product. The conceptual links that are missing as a result of increasing the success criterion are as follows:

1. Given two linearly independent vectors span a plane, and given three linearly independent vectors, there is no nontrivial linear combination that forms the zero vector.

2. The projection of vectors compares the direction of vectors; consequently, the projection of perpendicular vectors is the zero vector
3. Matrices are invertible if their determinant is nonzero.

Increasing the success criterion again from 50% to 75% results in the calculus sequence being almost completely eliminated from the concept map, as were the majority of linear algebra concepts. This means that *less than 75% of respondents* were able to provide evidence that they knew basic concepts, such as the Riemann Sum being computed as the sum of values a function takes on over a partition. The only concepts that were preserved at the 75% success criterion level are below:

1. The standard basis vectors in $\mathbb{R}^3$ are unit vectors.
2. If the dot product is zero, then the vectors in the product are perpendicular
3. Scalar multiplication scales the value of each vector component and preserves direction if positive.
4. The domain of integration for a two variable function can be determined from the endpoints of integration.

The maps imply that though the majority of students' misconceptions surrounding these structures come from the calculus course sequence. The prerequisite courses for enrollment in Math 3B03 contain three courses in calculus, and the calculus concepts addressed in this sequence build sequentially on concepts taught in grade twelve. Comparatively, there are only two linear algebra courses, and while there is some treatment of vector algebra in MCV4U, the focus of that course is differential calculus. Though the achievement at the 75% criterion is admittedly modest for linear algebra as well, it does suggest that the course sequence is comparatively stronger than the calculus sequence.

While we did expect the majority of items to have difficulty index between 0.4 and 0.6, we also hypothesized that there would be some concepts that were fundamentally understood by the majority of students. Consequently, we were surprised that none of the items would have difficulty index above 0.6. The items we used were intended to elicit thought about the relationship between concepts. In order to accomplish this, we needed to develop items that were not necessarily routine for participants. However, the items were largely straightforward, especially considering that they were testing concepts from Level I and II, and that students were given approximately a week to complete the inventory.

Category Analysis:

Though the concept map allows us to see an overview of the concepts that students miss in the inventory, the advantage of including item analysis in our study is that it gives is a tool to analyze what errors students are most commonly making with respect to the propositions being measured. To do this, we will analyze the distractors in categories (1), (3) and (9), since these fall in extreme ranges of difficulty and discrimination.

From this inventory, we see that there are two items in category (1), three items in category (3) and no items in category (9)[8]. By examining where items fall on a spectrum of difficulty and discriminatory power, we can examine which cognitive structures are commonly underdeveloped by students on the whole, and those that are only obtained by high achieving students. Lastly, the inventory was designed to include distractor items that could be feasible when using reasoning that is founded on conceptual errors. Analyzing the choices that students made in these extreme categories will tell us what conceptual deficiencies separate stronger and weaker students.

Category (1): Highly Discriminating and High Difficulty

Items category (1) have a discrimination index of at least 0.4, and a difficulty index of less than less than 0.4. These items were both highly correlated with the aggregate test score, and items that were frequently answered incorrectly; as such, we can interpret these items to be questions that were only accessible to students with strong cognitive structures of the content. From these questions, we can identify which concepts that are rarely mastered by students as a whole, but frequently mastered by high achieving students.

The two items come from the calculus course sequence. Since there are no terminological issues with either item, and each of the items' propositions were validated in cognitive labs and peer review, we can be confident that selections of each of the distractors effectively represent student reasoning through these problems. Thematically, it appears that the distinguishing feature between students that are successful on these items and those that are not suggests that students that are successful are able to integrate the algebraic representation of an object and its graphical meaning when they are presented in nonroutine settings.

[8] All items in the relevant categories are shown in Appendix D.

The first item in this category tests a student's ability to interpret a Riemann Sum. In particular, the item tests the students' ability to connect a Riemann Sum to its representation, shown below.

Example 5:

Consider the Riemann Sum of $x^2 - 1$ over the interval $[a, b]$:

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left(\left(-2 + \frac{4k}{n}\right)^2 - 1\right) \times \frac{4}{n}$$

Which of the following are true statements?

a. The limit of the sum represents the signed area bounded by $x^2 - 1$ over the interval $[a, b]$.
b. The limit of the sum represents the definite integral of $x^2 - 1$ over the interval $[a, b]$.
c. The Riemann Sum is computed over the interval $[-2,2]$.
d. (b) and (c)
e. (a), (b) and (c)

	A	**B**	**C**	**D**	**E***	**Total Responses**
Number of Responses	15	28	3	26	35	107

If successful on the item, a student will have provided evidence that they recognize the relationship between a Riemann Sum and the definite integral, are able to identify the geometric significance of the Riemann Sum as representing signed area of a function over an interval and can identify the region a Riemann Sum is being computed over within the representation of the sum. We found that while most students were able to recognize that the sum given did represent the definite integral of $x^2 - 1$ over an interval $[a, b]$, they were unable to connect the Riemann Sum to the signed area over an interval $[a, b]$ and to identify the interval as $[-2,2]$. This is surprising; students are introduced to Riemann Sums early in their academic career, shown diagrams of left and right sums, and are expected to reproduce them for assessment. Students are also asked to use Riemann Sums to approximate area under a curve, (and later, volume, length and surface area), and to approximate a definite integral over a closed interval. However, this concept is not effectively retained; or at the very least, the idea that the Riemann Sum is related to area under a curve it not effectively retained.

Perhaps most surprisingly, an item testing the knowledge of critical points, interpretation of the value of a derivative and the chain rule was also in this category. Students have been confronted with these concepts several times in their academic career, starting with their MCV4U credit, and are continually required use understanding of these concepts throughout their first- and second-year calculus sequence. The original version of this item proved challenging for students in the cognitive labs as well; however, we obtained substantial evidence that the option was testing what it was intending to test, further simplified the item, and the revision passed a subsequent peer review.

The item itself is seen in *Example 3*. It presents students with the graphs of two derivatives of positive functions. Students are asked to draw conclusions about each of the individual functions $f(x)$ and $g(x)$, and the derivative of the function $g(f(x))$. 95 of the 110 students who answered this question correctly identified that the root of the derivative corresponds to a critical point of the original function, and 75 respondents were able to successfully compute the sign of $g(f(x))$ within a given region; however, the distractor elements in the question indicated that 34 of respondents thought that because $g'(x)$ was monotone increasing, that $g(x)$ was monotone increasing, despite $g'(x)$ showing negative values in the graph.

The common feature between items in this category is the ability to translate between an algebraic and graphical representation of a mathematical object. In *Example 5*, the distractors indicate that most students failed to see equivalence between signed area, the definite integral and the Riemann Sum. Similarly, distractor selection in *Example 3* shows that students commonly misinterpret the graphical information presented to them on the first derivative and its relationship to the original function. Since items in this category represent items that were primarily achieved by high scoring students, we can see that students who excelled in the inventory distinguish themselves from lower-scoring peers based on the flexibility that they have when representing concepts in their head.

Category (3): Ineffectively Discriminating and High Difficulty

On the opposite side of the discriminating spectrum, we examine items in category (3). The items in this category have high difficulty index, but low discriminating power; this means that they are frequently missed by students with high and low total inventory scores, alike. Therefore, items in this category represent concepts that are rarely understood by students, regardless of their success on the inventory as a whole. All three of the items in this category are from the calculus sequence, involving properties of monotone functions, the definite integral and linear approximations. Unlike questions in category (1), we find that these items are related not by how students perceive relationships between of concept representations, but rather the ability to articulate properties of the objects described.

For instance, among items in category (3) is the following question, which assesses whether a student can identify a monotone function as being one to one. The modal response for this item was correct, followed by the selection that monotonicity was insufficient information to determine anything about the function. We would expect a Level III student to be able to reason that a function that always increases cannot take on the same value twice and construct a counter example for a monotone function that is not onto. We would certainly expect high achieving students to be able to do so—however, the low discrimination index for this item indicates that this item was challenging for students with and without high inventory scores, implying that both high and low achieving students seem to have difficulty articulating this property of monotone functions.

Example 6:

Suppose $f: \mathbb{R} \to \mathbb{R}$ is a differentiable function so that $f'(x) > 0$ for all $x \in \mathbb{R}$. Which of the following must be true for the function $f(x)$?

a. $f(x)$ is onto.
b. $f(x)$ is one to one.
c. $f(x)$ is either one to one or onto, but not both.
d. $f(x)$ is both one to one and onto.
e. Not enough information.

	A	**B***	**C**	**D**	**E**	**Total Responses**
Number of Responses	15	42	4	16	32	109

Similarly, the third item in the category asks students to identify true statements about a given linearization. The three sub-goals of this question are to assess whether students are able to articulate that a linearization of a two variable function is a plane, translate between the coefficients of a linear approximation and the tangent vectors of a function at a point, and distinguish between the value of a function and an approximate value given by a linearization. Based on the distractor selections from students, it appears students have difficulty with the third concept.

Example 7:

Suppose that the linear approximation of the function $f(x,y)\colon \mathbb{R}^2 \to \mathbb{R}$ at $(4,5)$ is given by:

$$L(x,y) = 12 + 3(x-4) + 2(y-5)$$

Which of the following is true?

a. The image of $L(x,y)$ is a plane.
b. The vectors tangent to $f(x,y)$ at $(4,5)$ are $\vec{\imath} + 3\vec{k}$ and $\vec{\jmath} + 2\vec{k}$.
c. The value of $f(5,5)$ is 15.
d. (a) and (b)
e. (a), (b) and](c)

	A	B	C	D*	E	Total Responses
Number of Responses	21	18	7	31	32	109

The most jarring misconception found in this category was seen in the item on the definite integral in response to the item in *Example 1*. Adapted from the *Good Questions* project (Terrell, n.d.), this item indicates that students are associate the definite integral over a region with the area under a curve despite this only being true for nonnegative functions. In fact, more students thought that the definite integral is an antiderivative than that it is a finite quantity. The summary of responses of this item are shown below.

	A	B*	C	D	E	Total Responses
Number of Responses	69	14	19	5	3	110

It's also worth noting that though students overwhelmingly believe that the definite integral is the are under a curve, they did not believe the same of the Riemann Sum in *Example 5*, even though the phrasing of *Example 5* used "signed area," which is correct. At the class level, this shows that students hold beliefs about integrals and Riemann Sums that are inconsistent with the definition of what these objects are. This goes along with the findings of the FCI and CCI (Epstein, 2013), showing that students can undergo instruction, but still hold beliefs that are logically inconsistent.

Addressing Misconceptions in Undergraduates:

The results of analyzing our first inventory suggest that students have many calculus and linear algebra related misconceptions when entering their upper-level undergraduate studies. Previous studies in undergraduate education in a variety of disciplines tell us that student misconceptions are not alleviated on their own; instead, students carry their misunderstandings with them, and try to fit newly learnt material into their faulty cognitive structures (Epstein, 2013). This makes addressing misconceptions challenging, since increased instruction in the subject area alone does not result in a meaningful change in misconceptions (Lai, 2009).

Altering instructional strategies from traditional lecture style to those that favour interactive engagement has been shown to improve student misconceptions and performance on both the FCI and CCI (Lai, 2009). These strategies include peer instruction methods, where students are presented with conceptual questions, and given between one and two minutes to prepare and report their answers to instructors. They are then given the opportunity to collaborate with their peers on the questions after seeing results of the poll for an additional two-minute period. Students report their final answers via a poll, and lecture continues. This method was used by Hake in the early 1990s, where he observed that students engaging in peer instruction methods results on the FCI doubled compared to those who were taught using traditional lecture styles (Lai, 2009).

In his study on the CCI, Epstein reported that students taught by instructors that predominantly use lecture style instruction methods experience a normalized gain of 0.23 on the calculus concept inventory over the course of a semester; in contrast, those that were taught using interactive engagement strategies experience a normalized gain of 0.48 (Epstein, 2013). This is in line with similar studies in other disciplines, where students who are taught using interactive engagement outperform their peers taught by traditional lecture styles in terms of conceptual understanding (Epstein, 2013).

As a result, there have been efforts in recent years to develop resources for interactive engagement in large calculus classes (Lai, 2009). This includes the *Good Questions* project, which provides instructors with a pre-made multiple choice and true/false questions designed to be used in peer instruction formats (Terrell, n.d.). The items are designed to expose conceptual subtleties and common misconceptions that students have while learning new material, categorized by chapters in James Stewart's *Calculus Concepts and Contexts Single Variable* textbook. Each question includes instructional suggestions and the misconceptions that may arise in discussion, making the resource extremely accessible for novice and experienced instructors alike (Terrell, n.d.). In a study that examines the effectiveness of using these questions, Terrell finds that using problems regularly substantially increased student performance on final exams compared to their preliminary exams (Lai, 2009).

A common counterargument for using this approach is that interactive engagement requires an investment of lecture time. This is a fair criticism, especially for the first and second-year courses ,where the amount of material to cover makes the opportunity cost of time high. However, it is worth considering how useful the coverage of material is when possibly only a small fraction of a class is meaningfully learning that material. It could also be worth revisiting the curricular goals of the first- and second-year sequences to reduce the coverage of material, affording more discretionary instructional time in the prerequisite course sequence.

Further Concept Inventories:

From a practical point of view, the later concept inventories do give us some indication that student understanding of content as they progress through the course, since they ask questions relevant to new course material as it is taught. We can therefore compare results from inventories in the course; however, the $KR20$ values of 0.217 for Inventory 2, and 0.519 for Inventory 3 do not provide sufficient evidence of internal consistency to draw any strong conclusions from their results based on classical test theory (Frankel & Wallen, 2009). This could be the case for a variety of reasons, including the smaller inventory size or survey fatigue; however, we believe that there may be another explanation that warrants further investigation.

In our first concept inventory, we aimed to measure cognitive structures that supposedly already existed for students upon entering the class. This is a standard procedure with concept inventories; both the FCI and CCI are administered well after students have learned the material they aim to assess. However, our later concept inventories aimed to measure the development of *new* structures, which has not been explored using concept inventories in this context. The theoretical underpinning of this is that if there are relatively unstable structures shown at the class level for the first inventory, then the success for concepts that are learned later on would also be unstable. In other words, that failure to establish a rich cognitive structure on prerequisite concepts would influence how successful students are in later concepts.

Little is known about how students learn mathematics, especially at the undergraduate level. The body of literature primarily consists of pre- and post-assessments, which aim to establish how students perform in comparison with some benchmark of achievement. However, these studies examine student understanding prior or after learning has already happened, rather than student understanding when learning is in progress. Administering the second and third concept inventory was our attempt at measuring the development of cognitive structures throughout the learning, as suggested by Castles and Lohani in their paper on modelling student understanding (Castles & Lohani, 2009), in order to explore how newly developing cognitive structure attaches itself to pre-existing ones.

When we consider that "pre-existing" and "developing" cognitive structures may be fundamentally different variables, it is not surprising that the internal consistency measures for our later concept inventories are substantially lower than the $KR20$ value for our first inventory. Low internal consistency may actually be what we should expect from an inventory result that is attempting to measure understanding as it develops, as novice students may not recognize and apply the intended concepts correctly.

Issues with internal consistency aside, the idea that developing and pre-existing cognitive structures are different constructs raises new questions for how each of these structures should be measured which has consequences for the model suggested Castles and Lohani in the context of educational research. Ausubel's Theory of Subsumption tells us that they are at the very least related, and that the stability of newly developed structures relies in part on the structural richness of pre-existing structures (Ausubel, 1963), and it seems that these two variables should be intimately related, from an epistemological point of view. For those reasons, using an auto-generated concept map from a concept inventory may be appropriate as a feedback mechanism for students throughout the learning process as an alternative to binary feedback in practice. However, demonstrating that there is a relationship between these constructs empirically in a scalable, measurable way remains an open question, which limits the utility of auto-generated concept maps of learning in progress for research purposes.

Addressing the Research Questions:

This project originated from a need to understand the extent to which students in the undergraduate program acquire necessary cognitive maps for mathematics as they progress through their prerequisite course sequences.

By examining the concept maps generated from the concept inventories, we can see that the calculus course sequence has ineffectively scaffolded the understanding for students in the sample compared to the more stable cognitive structures from the linear algebra sequence. Using techniques in item analysis, we can see that stronger students appear to have an enhanced ability to integrate algebraic and geometric representations of mathematical objects compared to their peers. We also see that there are persistent misconceptions among undergraduates for the calculus sequence that are inconsistent with material being taught in courses. It appears that the course sequence for calculus is ineffectively preparing students for upper-level undergraduate work, and that the linear algebra sequence has been comparatively more successful in preparing students for higher level studies in differential geometry.

This effectively answers our first research question; that is, that the calculus sequence is ineffective in establishing stable cognitive structures for differential geometry compared to the linear algebra sequence. While this may be surprising, it is worth noting that this relationship—or lack thereof—between conceptual understanding and time spent in university courses is consistent with Epstein's findings with the CCI, where a semesters' instruction in calculus did not result in a meaningful change in students' conceptual understanding of the discipline (Epstein, 2013). Though Epstein's methodology examined normalized gains over the course of a semester on the same inventory, and primarily investigated one semester of first year calculus, the findings are relevant here since they show that instruction has not effectively developed calculus understanding for the students in our sample. The FCI findings are similar; instruction does not in and of itself result in strong student understanding (Epstein, 2013).

This shows that even though students have more experience with calculus related content, they do not have comparatively stronger conceptual frameworks of the material compared to linear algebra, which is reasonably novel to the students and taught in fewer courses. In other words, the required courses did not result in a meaningful ability to reason with concepts taught in calculus, especially compared to those in linear algebra.

One possible explanation for the difference in curricular expectations for the two course sequences. Due to large class sizes and content required to be taught in the calculus course sequence, many of the first- and second-year courses use traditional assessments involving primarily computational answers. In theory, students could pass through these courses having learned how to effectively compute answers to questions, but not having learned what the answers they computed mean in the context of calculus.

In contrast, the linear algebra sequence requires students to demonstrate and "explain some theoretical underpinnings" of the discipline in their first course in linear algebra. MATH 2R03 is a primarily proofs-based course, that relies heavily on a strong understanding of theory. These assessments require students to be thoughtful about relationships between concepts and apply them effectively in order to receive credit in the course. While it may also be that the linear algebra material necessary for differential geometry was admittedly more rudimentary than the concepts required for other branches of mathematics (for example, that necessary for abstract algebra), the cognitive labs show that items developed did test concepts they were intended to at a sufficiently high level to be confident that they accurately reflect second year material.

The second aim of the study was to establish whether there was a measurable, predictive relationship between cognitive structures and academic success. In theory, it is reasonable to think that there would be a predictive relationship between cognitive structure and academic achievement. Ausubel's theory of subsumption supports the idea that students who have rich structures are capable of enhanced retention of new material and allows students to make connections between concepts (Ausubel, 1963).

Though we did see correlation between the first inventory scores and the aggregate test scores over a semester, the sample size is unfortunately too small to draw any conclusions. Furthermore, though we did attempt to measure later cognitive structures as they develop throughout the course and found relatively strong correlation between those inventories and test scores, the Cronbach Alpha coefficients for those inventories did not give sufficient evidence of internal consistency. For those reasons, the strong correlation coefficient is encouraging that there may be a relationship; however, replication of the study is needed in order to be confident in the predictive power of the inventory developed.

Conclusion:

We began this study hoping to better understand cognitive structures of the students entering upper-level undergraduate studies in mathematics at McMaster University, and assess their current understanding of the fundamental concepts taught in their first two years of study. After examining the results of our concept inventory in the context of relevant literature, we have produced evidence of variances in how effectively the prerequisite courses in calculus and linear algebra develop students' conceptual understanding of the discipline. Our findings contribute meaningfully to the discussion on undergraduate mathematics curriculum at McMaster, and propose actionable, practical suggestions for improving the conceptual understanding of mathematics for our undergraduates. Of course, our research has implications beyond local, i.e., beyond informing curricular changes at McMaster.

In our study, we successfully developed a concept inventory that is internally consistent, has content validity, and is linked to a graphical depiction of a students' cognitive structure of introductory material. This is the first inventory of its kind that attempts to measure conceptual understanding of cognitive structures relevant for upper-level mathematics. By examining the concept domain for the discipline, we are able to discern what concepts are most relevant for students studying differential geometry and measure the extent to which they have mastered those concepts. Further, we have developed a method for representing cognitive structures using a concept inventory as suggested by Castles and Lohani (Castles & Lohani, 2009), and preserve the logistical and analytic advantages that concept inventories afford using classical test theory (Bai & Ola, 2017).

Replication of the study is needed to ensure that the findings on the internal consistency of the instrument hold true on a more representative sample, with varying institutional contexts. Should the results of future studies be consistent with this one, we will have created a concept inventory that is appropriate for large-scale research. This is a significant achievement; there are few conceptual instruments that are suitable for large-scale studies, a considerable barrier for research in the field. This inventory has the potential to enable the development and administration of large-scale research studies, opening new avenues for future research on students' conceptual learning.

The study has also exposed new areas that warrant further investigation. Having observed a positive correlation between the prerequisite concept inventory and test achievement of differential geometry students, replication would allow us to further explore whether this relationship holds true more generally, or if this relationship was sample dependent, or course (content) specific. Research on the predictive power of the inventory on student academic outcomes could motivate the development of similar inventories for gate-keeper courses in the program, such as Real Analysis. Furthermore, our subsequent concept inventories and cognitive labs motivate theoretical questions on the development of conceptual understanding of mathematics.

Our work has shown that the development of mathematical understanding is a nontrivial process, that requires considerable curricular and instructional attention. We believe in the potential our undergraduates have to learn and appreciate rich mathematical ideas, and hope that this study inspires an effort among instructors to intentionally facilitate concept-oriented learning experiences to assist in developing concept-knowledge.

Bibliography

1. Ausubel, D. (1963). Cognitive Structure and the Facilitation of Meaningful Verbal Learning. *Journal of Teacher Education, 14*, 217-222.
2. Bai, X., & Ola, A. (2017). A Tool For Performing Item Analysis to Enhance Teaching and Learning Experiences. *Issues in Information Systems, 18*(1), 128-136.
3. Castles, R., & Lohani, V. (2009, 01 01). A paradigm for comprehensive conept map-based modeling of student knowledge. *ASEE Annual Conderence and Exposition, Conference Proceedings*.
4. Epstein, J. (2013, September 2013). The Calculus Inventory -- Measurement of Effective Teaching Methodology in Mathematics. *Notices of the AMS, 60*(8), 1018-1026.
5. Frankel, J. R., & Wallen, N. E. (2009). *How to Design and Evaluate Research in Education.* New York: McGraw Hill.
6. Furrow, R. E., & Hsu, J. L. (2019, 01 16). Concept inventories as a resource for teaching evolution. *Evolution: Education and Outreach, 12*(1).
7. Gleason, J., White, D., Thomas, M., Bagley, S., & Rice, L. (2015). The calculus concept inventory: A psychometric analysis and framework for a new instrument. *19th Annual Conference on Research in Undergraduate Mathematics Education*, (pp. 135-149).
8. Institute for Machine and Human Cognition. (2003). *A Summary of Literature Pertaining to the Use of Concept Mapping Techniques and Technologies for Education and Performance Support.* Pensacola: The Institute for Human and Machine Cognition.
9. INSTITUTIONAL RESEARCH AND ANALYSIS, MCMASTER UNIVERSITY . (2020). *McMaster University Factbook 2019|2020.* Hamilton: INSTITUTIONAL RESEARCH AND ANALYSIS, MCMASTER UNIVERSITY .
10. Ivie, S. D. (1998). Ausubel's Learning Theory; An Approach to Teaching Higher Order Thinking Skills. *The High School Journal, 82*(1), 35-42.
11. Jorian, N., Gane, B. D., DiBello, L., & Pellegrino, J. W. (2015, June). *Developing and Validating a Concept Inventory*. Seattle, Washington, United States of America.
12. Lai, Y. (2009, May). Teaching Undergraduate Mathematics. *Critical Issues in Mathematics Education, 5*, 37-48.
13. Lucidchart. (2020). Retrieved from Lucidchart: https://www.lucidchart.com/
14. Madsen, A., McKagan, S. B., & Sayre, E. C. (2017, December). Best Practices for Administering Concept Inventories. *The Physics Teacher, 55*(9).
15. *Mathematics Stack exchange*. (2020). Retrieved from StackExchange: https://math.stackexchange.com/

16. Melhuish, K. M. (2015). The Design and Validation of a Group Theory Concept Inventory. Portland, Oregan, United States of America: Portland State Univerity.

17. Mintzes, J. J., Markham, K. M., & Jones, G. (1994). The Concept Map as a Research and Evaluation Tool: Further Evidence of Validity. *Journal of Research in Science Teaching, 31*(1), 91-101.

18. Pressley, A. (2012). *Elementary Differential Geometry.* Springer.

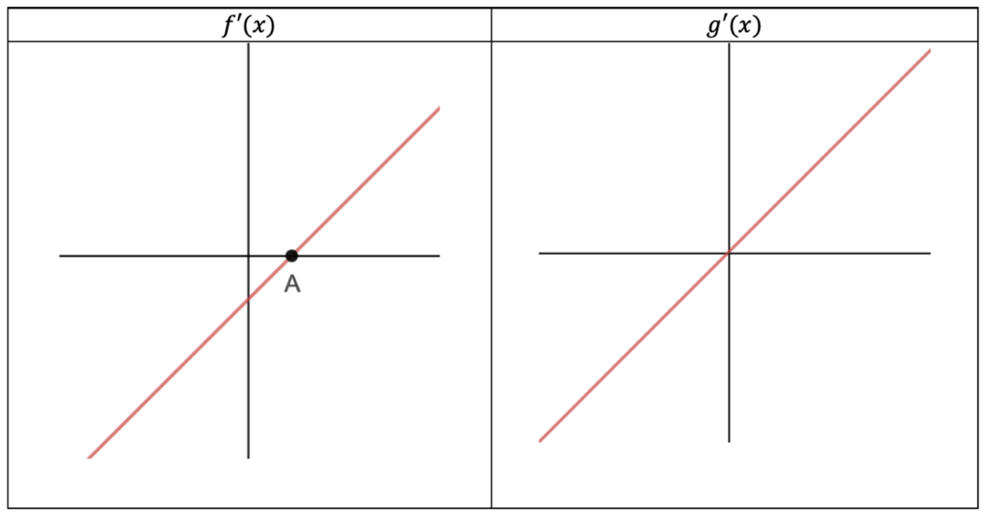
19. Raussen, M. (2008). *Elementary Differential Geometry: Curves and Surfaces.*

20. Terrell, M. (n.d.). *Class Materials*. Retrieved June 2020, from GoodQuestions Project: http://pi.math.cornell.edu/~GoodQuestions/materials.html

Appendix A: Sample Validity Tests from Cognitive Labs

Question 4

Suppose that $f: \mathbb{R} \to \mathbb{R}$ and $g: \mathbb{R} \to \mathbb{R}$ are positive for all $x \in \mathbb{R}$, differentiable functions, and the graphs of their derivatives are given below:



What can we conclude about the following composite functions?

a. $f(f(x))$ is increasing for all $x \neq A$
b. $g(f(x))$ is has critical points at $x = A$ and $x = 0$
c. $f(g(x))$ is decreasing for $0 < x < A$
d. $g(f(x))$ is increasing for $x < 0$
e. $f(g(x))$ has no critical points

Option	Desired Conclusion	1	2	3	4	Latent evidence for Validity of Conclusion	Manifest Evidence for Validity of Conclusion
A							2: So, we get $f'\big(f(x)\big) \times f'(x)$

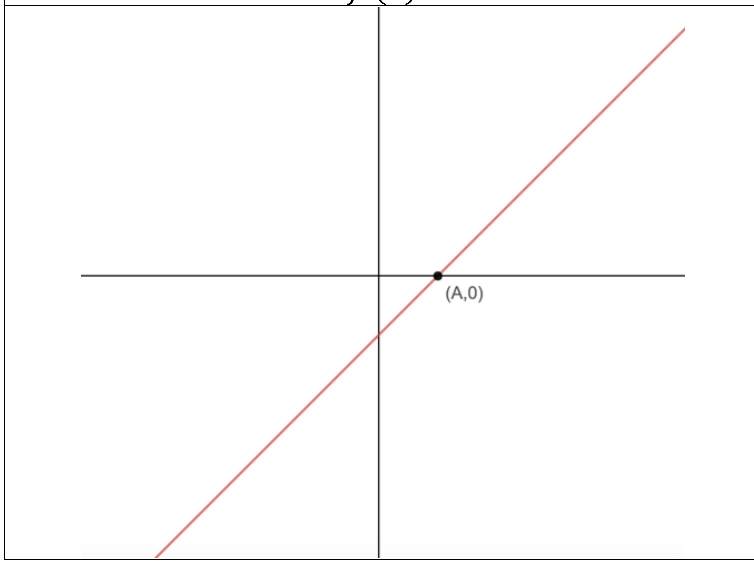
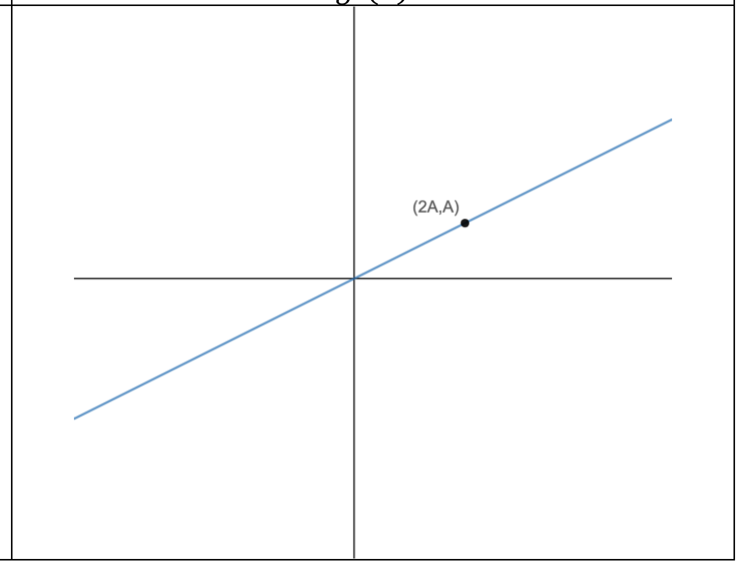
B	The derivative of composite functions is proportional to the derivative of the outer function						2: F of g…its derivative would be $f'(g(x)) \times g'(x)$. We want to know when this is equal to zero. See that's the thing, we don't know anything about $g(A)$. So, I would not pick B. 3: I'm tempted to say B because there would be two critical points because of the fact that g and f have similar derivatives.
C						3: This thing is increasing entirely, so D kind of makes no sense, and C makes even less sense. 1: Like if you have g'(x) and you move it to f'(x), then it would be decreasing, because you would see this line moving downwards. So, so then the line would intersect. It would go from intersecting 0 to intersecting A, so it's going down, so its decreasing.	2: Between 0 and A, that's positive, so g' will be positive, and then we see that…definitely 'c'. I: Okay, why is that? 2: Because $f'(g)$ is going to be negative, and g is going to be positive, and we know that for any positive…oh, actually. Never mind. I don't like that." 1: So, if f(g(x)) is equal to, actually right, I would check the derivative so that would be a product rule. No, not the product rule, a chain rule. So that would be f'(g(x))*g'(x)=0.
D						3: This thing is increasing entirely, so D kind of makes no sense, and C makes even less sense.	
E							

General Comments	1: So, if f(g(x)) is equal to, actually right, I would check the derivative so that would be a product rule. No, not the product rule, a chain rule. So that would be f'(g(x))*g'(x)=0. 1: These derivatives seem to have the same slope…f(g(x)) is decreasing, because, well…the input is g(x) to f(x) and the slope of f'(x) looks like it's the same but its shifting… 1: Well, I don't know what the graph of g(x) looks like though. 2: "We know that f has a minimum…is it a minimum? Yes, it's a minimum at A." "Oh, well. I could use the chain rule" "See that's the thing, we don't know anything about $g(A)$." "I have nothing that tells me that $g(x)$ is going to be between 0 and A, so I don't know where $g(x)$ is between 0 and A. All I know is that it is positive, and that doesn't prevent f from being positive. I was basically going to have f' is negative, and then I would have had negative times positive is negative." 3: "I would stick with B, simply because there are two zeros and when you compose two functions, most of the time, a lot of their properties get combined." 4: "I don't think I know this one. I: Okay. What about it is making you think you don't know it? Are there specific words, or aspects of this that is throwing you off?" 4: Uh…no, it's primarily memory of the subject. This is all the way back to first year first semester."

Conclusion: REVISE AND INCLUDE IN PEER REVIEW	This question is definitely eliciting the kind of thinking it is intended to; for instance, both participants that answered correctly both referenced the chain rule and computed it effectively, which gives solid evidence that this structure of question involving graphs of functions derivatives and asking about compositions of derivatives. While this does require computation on part of the student, it doesn't involve direct computational reasoning, which makes it an appropriate style to test the chain rule in accordance with the purpose of this study. That said, it is not reasonable to say that this question is assessing what it is intended to assess as it is currently written. This is partially due to the fact that it involves so many aspects of calculus, such as interpreting the derivative, positive or negative derivative resulting in increasing or decreasing functions, a critical point of a function being the zero of a derivative, and derivative rules. Though this would definitely make an excellent test question, it is not currently appropriate to assess the properties of the chain rule, as it tries to test too many aspects of derivatives and properties of functions implicitly as opposed to explicitly. In order to improve the quality of the inferences gained from this question, it needs to test these properties explicitly. Over and above that, there are issues in the question itself, such as not providing sufficient information about the values of $g(x)$, and the slopes of both of these functions being equivalent which led participants into answer the correct answer (c) by using invalid reasoning. Therefore, the question needs to be rewritten, and re-evaluated in peer review. There was also an interesting comment made by participant 3, that gets at another interesting misconception which may be beneficial to include: "This thing is increasing entirely, so D kind of makes no sense, and c makes even less sense" From the transcript, it appears as though this participant believes either that because the values of the function is always positive and the slope of the derivative is always positive that the function must always be increasing, or that because the slope of the derivative being positive implies that the function is always increasing. As such, it may be beneficial to include an option that lends itself to this kind of thinking to see if it is a prevalent misconception. In an attempt to preserve the aspects of the question that were effective in eliciting student thinking, the format of the question will still present two graphs of derivatives and ask about the derivative of a composite function. However, there will also be options that indicate increasing/decreasing and critical points of the individual functions to explicitly test these behaviours on their own. This will allow the conclusions being reached by each option to explicitly tested.

A suggested modification to this question is as follows:

Suppose that $f: \mathbb{R} \to \mathbb{R}$ and $g: \mathbb{R} \to \mathbb{R}$ are differentiable functions, and the graphs of their derivatives are shown below:

$f'(x)$	$g'(x)$
	

If $f(x) > 0$ and $g(x) > 0$ for all $x \in \mathbb{R}$, then which of the following are true?

f. The function $g(x)$ is increasing for all values of x
g. The function $f(x)$ has a critical point at $x = A$
h. The derivative of $g(f(x))$ is negative for $0 < x < A$
i. (b) and (c) only
j. All of the above

With these modifications, the propositions that are assessed in this question can be altered to the following:
A: If the derivative is positive, then the function is always increasing.
B: If the derivative is zero at a point and not zero on either side, then the function has a critical point.

	C: Given information about f, g, f' and g', then we can conclude information about the derivative of $f(g(x))$ using the chain rule. D: Both conclusions in B and C E: All three conclusions.

Question 8

Suppose that $f(x)$ is continuous and integrable over $[a, b]$.
Which of the following is always true?

a. $\int_a^b f(x)dx$ is the area bounded by the graph of f, the x-axis and the lines $x = a$ and $x = b$
b. $\int_a^b f(x)dx$ is finite
c. $\int_a^b f(x)dx$ is an antiderivative of $f(x)$
d. $\int_a^b f(x)dx$ may not exist
e. None of these

Option	Desired Conclusion	1	2	3	4	Latent evidence for Validity of Conclusion	Manifest Evidence for Validity of Conclusion
A	The definite integral of an integrable function represents the area bound by the x axis and the curve over a given interval					3: Well (a) is wrong because it's not necessarily bounded by the x axis depending on what the graph looks like. I: Alright, and what exactly do you mean by that? What do you mean by "not necessarily bounded"?	2: I'm a little iffy about the wording of A […] because the definite area is not just the area bounded by those things, it's the signed area. So, it's the area when you take into account the signs. 3: The integral is the net area bounded by the graph.

							4: The area bounded by the graph of f. Yeah, it's a. That's the first definition of integration, if I'm correct.
B	The definite integral of an integrable function is necessarily finite.					3: Why is one of the questions that f(x) is finite? […] If f(x) is continuous and integrable over an interval […] it being finite seems like an obvious always true answer.	"2: I'm tempted to say B [...] well we have continuity and we have integrability. I think it is B. I: And that's because of what statement in the question? 2: Really, it's because of continuity." 3: […] I'd say that its finite because normally it's a quantity you can calculate, whether there are methods to calculate it or not. 4: Option B might be true if the graph goes to infinity at b or a. Technically there would be no finite area then, even though it is integrable.
C	The definite integral of an integrable function is the antiderivative of the function.						2: C is not true the way it's written…that's kind of weird. Because an antiderivative, maybe it's the right idea, but it's definitely not the right way of writing it. An antiderivative is something that when you take the derivative of it you get the function. That right there is

							a definite integral, so it's going to give you a number. 3: If you have a definite integral it's not necessarily an antiderivative, its more of a numerical quantity, whether you can calculate it or not. 1: Would the integral—would a definite integral ever not be an antiderivative? I don't think so, since that's the fundamental theorem of calculus.
D	The definite integral of an integrable function may not exist						2: Well integrability gives that its integrable so D is not true. 3: If f(x) is continuous and integrable, then D is wrong.
E	None of these.						
General Comments	1: I was thinking that if…well, it says its bounded by some area, but then I was thinking if you could substitute one of the values for something like infinity or negative infinity, so I backpaddled out of thinking that.						
Conclusion: VALID	Based on the manifest evidence provided by the participants, each option is effectively targeting the misconceptions and conceptual frameworks it is intended to target. As such, this question is passed to the inventory without requiring further peer review.						

Question 10

Consider the following Riemann Sum:

$$\lim_{n\to\infty}\sum_{i=1}^{n}\left(\left(-2+\frac{4i}{n}\right)^2-1\right)\times\frac{4}{n}$$

Which of the following is true?

a. The limit of the sum represents the area bound by $x^2 - 1$ over some interval $[a, b]$
b. The limit of the sum represents the definite integral of $x^2 - 1$ over some interval $[a, b]$
c. The Riemann Sum is computed over the interval $[-2,2]$
d. (b) and (c)
e. (a), (b) and (c)

Option	Desired Conclusion	1	2	3	4	Latent evidence for Validity of Conclusion	Manifest Evidence for Validity of Conclusion
A	The limit of a Riemann Sum always represents the area of the function.						2: I'm not super happy with (a). I like it when they say signed area, so this is making me a bit uncomfortable. 2: I picked (a) because I don't want to be pedantic.
B	The limit of a Riemann Sum represents the definite integral of the function.					2: Right, so I'm basically trying to re-derive the definition of the integral.	2; I settled on B and C because those were things I could actually verify. [..] it is a definite integral by definition
C	The endpoints of the Riemann Sum can be determined by					2: The thing that is bothering me now is the way the terms	2: Now this is a little annoying. Yeah, by minus 2 plus 4i over

	examining the input to the function within the Riemann Sum.					are stacked. Because…F at x i times delta x.	n? I think there’s some manipulation here that I have to be doing. […] Basically, this thing is counting the thing inside the bracket. Its starting with -2, and its adding multiples of 4/n, which basically takes you to the next rectangle, next rectangle, next rectangle. So I’m relatively sure that B and C are correct. 2: First of all, the intervals length is 4, or it should be 4, and the starting point is -2. 3: I don’t know where you’re getting this function from, or even the interval.
D	B and C						
E	A, B and C.		■				2: Okay, I’ll say E. then. […]
General Comments	1: I don’t think we dealt with Riemann sums… 1: I suppose like, if I took the limit as n approaches infinity, it would be…well, it’s not a geometric sum. 1: ‘i’ kind of throws a wrench in things because I’m not sure how that operates as a constant. 1: I’m trying to approximate it. 2: Oh. I thought i was an imaginary number. 2: You take this graph, which is possibly x squared minus 1. Then you break it up over a, b, and you add up the rectangles. Then you take the sum, so… “I: And you know it’s x squared minus 1 because? 2: Well, I don’t know for sure. I’m sure if I thought about it more something else could pop up. But also, the way that the choices are set up, are influencing my decision as well.”						

	3: I actually don't know the answer to this question. I'm serious. I remember Riemann sums, none of this is even beginning to make any sense. 4: I really wish I remember how to do this, but I don't think I do. I remember I studied this for the exam, and then I never touched it again.
Conclusion: VALID	Only two of four participants attempted this question, of which only one settled on an answer. Bearing that in mind, the fact that participants were not able to answer this question does not in and of itself disqualify this question from being valid. The idea of Riemann Sums is critical to understanding the development of material in differential geometry, and the construction of a Riemann Sum is—or should be—extensively treated in the undergraduate curriculum. Over and above that, participants that struggled with this question did so for easily documented reasons that point to the validity of the options. For instance, participant 3 referred to not knowing where the endpoints or the function could come from in the question. Furthermore, the participant that did attempt this question gave very explicit confirmation that each option was effectively testing what it was intending to test. The participant gave valid explanation for their choice of both B and C, and while the participant did wrongly select the option A, they did so because they didn't "want to be pedantic"; as such, the options presented here and the phraseology in the statement of the question are valid. Therefore, the question can be passed to the inventory as is, with minor modifications outlined below: Consider the following Riemann Sum of $x^2 - 1$ over the interval $[a, b]$: $$\lim_{n\to\infty}\sum_{k=1}^{n}(\left(-2+\frac{4k}{n}\right)^2 - 1)\times\frac{4}{n}$$ Which of the following are true statements? a. The limit of the sum represents the signed area bound by $x^2 - 1$ over the interval $[a, b]$ b. The limit of the sum represents the definite integral of $x^2 - 1$ over the interval $[a, b]$ c. The Riemann Sum is computed over the interval $[-2,2]$ d. (b) and (c) e. (a), (b) and (c)

Modifications include changing the index of the sum i to the index k, since two participants found this to be initially confusing and the interpretation of the index variable as an index rather than a complex number is not relevant to the propositions being assessed. Additionally, the integrating function $x^2 - 1$ was added to the question, as the options already were giving away that the function was $x^2 - 1$ as evidence by a comment made by participant 2. Lastly, the phrasing of option (a) has also been altered to say, “signed area”, which will now alter the correct answer be (e). The reason for this is that the concept that the definite integral and Riemann Sum does not always equal the area is being effectively tested in other questions already, so adding the term “signed” will allow students in the Phase II to indicate that they can distinguish between these topics.

Based on the modifications outlined above, the revised propositions are below:

A: The limit of the Riemann Sum represents the signed area bounded by the given function over the interval.
B: The limit of the Riemann Sum represents the definite integral of the given function over the interval.
C: The endpoints of the Riemann Sum can be determined by examining the input to the function within the Riemann Sum.
D: B and C
E: A, B and C

Because these are all minor changes, this question is passed to the inventory without requiring further peer review.

Question 12

Which of the following are unit vectors?

a. $v = 0.5i + 0.5j$
b. $\frac{v}{|v|}$, where v is the vector in (a)
c. i
d. (a), (b) and (c)
e. (b) and (c)

Option	Desired Conclusion	1	2	3	4	Latent evidence for Validity of Conclusion	Manifest Evidence for Validity of Conclusion

A	A unit vector has component magnitudes that sum to 1.					*Participant 4 attempted to compute the magnitude of this vector. Based on their description of their computations, it is clear that there was a calculation error that prevented them from seeing that A was not a unit vector.*	1: I just need to compute their length and see if it's one. […] the square root of one half isn't 1. 3: (a) after realizing its Euclidean, the length would also be equal to one, which is the definition of a unit vector.
B	A vector divided by its magnitude is a unit vector.					4: Okay, then it should be b.	1: Well actually, v over magnitude of v, that is a unit vector. […] well, B would be true because you're normalizing the vector. 3: (b) is the definition of the unit vector that I remember.
C	The standard basis elements for $\mathbb{R}^3$ are unit vectors					2: I'm going to assume it's one of the standard vectors. Okay, then B and C.	1: Well yeah, i is a basis vector, so it would be a unit…well, because it's not multiplied by anything. 3: No, I remember that i and j are the unit vectors in…[…] (c) because it's an example of a unit vector. 4: The magnitude of this vector I is one.
D	A, B and C						
E	B and C.						
General Comments	1: Does it mean the 'i' component of v? 3: Because if you half the i and j components of the vector then you don't know how long those components are in the first place. 4: v over mod v prime…I don't know what mod v prime is. 2: If i means the imaginary….okay, well that doesn't matter. I'm going to assume it's one of the standard vectors.						

	4: Okay. So, i… is I the universal mathematical vector?
Conclusion:	This question appears to be valid based on the evidence presented. It is passed to the concept inventory as it is written, with the following notational changes: Which of the following are unit vectors? a. $v = 0.5\vec{\boldsymbol{i}} + 0.5\vec{\boldsymbol{j}}$ b. $\frac{v}{\lvert v \rvert}$ where v is the vector in (a) **c.** $\vec{\boldsymbol{i}}$ d. (a), (b) and (c) e. (b) and (c) The propositions being assessed are the same as in the original question.

Question 14

The dot product of two vectors u, v is positive. Then we can conclude:

a. The vectors are pointing in the same direction
b. At least one vector has positive components
c. The angle contained between u and v is acute
d. The angle contained between u and v is obtuse
e. The vectors are linearly dependent

Option	Desired Conclusion	1	2	3	4	Latent evidence for Validity of Conclusion	Manifest Evidence for Validity of Conclusion
A	If the dot product is positive, then the vectors are pointing in the same direction.					"2: If you take a vector and dot it with itself, it does make	1: Well, no…you can have positive vectors that aren't

						a positive number, but that doesn't imply the converse."	necessarily pointing in the same direction.
B	If the dot product is positive, then at least one vector has positive components.						1: If you take the dot product, then you're basically multiplying the components together and adding them, so you can have two negative components together and adding them, so you can have two negative vectors, like two vectors with negative components, but their dot product would be positive […] 2: At least one vector has positive components, well you don't know that because you just take the magnitude… 4: That need not be, because you can apply two negative components and get a positive. So, it's not B.
C	If the dot product is positive, then the angle between the vectors is acute.					1: I don't think we can determine things about the angle, at least as far as I can tell, I'm not sure how you would do that with just the fact that it's positive.	2: The dot product of u dot v is equal to the magnitude of u, magnitude of v cosine theta, and if it is positive, it means cosine theta is positive, so you have to be either in the first or fourth quadrant. So, the angle between them should be acute? Yeah, should be acute. 3: When cosine is between a certain range of angles, between 0 and pi by 2…the

			■	■	■		angle is acute. And since the dot product of the two vectors can only be positive or negative because of the angle between the vectors, especially when you use the formula, which is equal to the norms of those two vectors, which has to be positive, times the cosine of the angle between those two vectors, I kind of realized between those two things its C. 4: I think C. I think its acute. Because cosine is positive in the first quadrant, and anything past that is negative until the fourth quadrant. So, if its past 90 its going to be a negative answer, and the dot product is positive, so it has to be in the first quadrant.
D	If the dot product is positive, then the angle between the vectors is obtuse.					1: I don't think we can determine things about the angle, at least as far as I can tell, I'm not sure how you would do that with just the fact that it's positive.	4: So, cos 90 is zero, and anything above cos 90, that's getting, wait…anything above 90 is getting negative. So, it's definitely not obtuse. So, it's not d.
E	If the dot product is positive, then the vectors are linearly dependent.	■				4: Why is the dot product being associated with linear dependence? I don't think that makes sense.	"I: Can you tell me why you're not picking E? 2: Basically, they could be linearly independent […] I mean I know an example when they are linearly dependent. If

						Participant 1 eliminated each of the other options and answered E. There was no particular conceptual reason for that choice stated by the participant other than the implicit process of elimination.	you take a vector and dot it with itself, it does make a positive number, but that doesn’t imply the converse.”
General Comments							
Conclusion: VALID	The responses to this question show substantial evidence that the question is testing what it is intended to test, so this question is passed to the concept inventory as written.						

Question 15

Suppose that we have two vectors $u, v \in \mathbb{R}^3$. If the cross product $v \times u$ is a nonzero vector w. Then:

a. w is perpendicular to u and v.
b. u, v span a plane.
c. u, w and v are linearly independent vectors.
d. (a) and (b)
e. (a), (b) and (c)

Option	Desired Conclusion	1	2	3	4	Latent evidence for Validity of Conclusion	Manifest Evidence for Validity of Conclusion
A	If cross product of two nonzero vectors is a nonzero vector, then it is perpendicular to the two vectors in the product.					1: Yes, I think so, because cross products, I believe…in differential geometry they use cross products for determining Frenet Frames? […] I believe it is true that	2: (a) is true. That’s just the definition of the cross product. 3: Oh, its E, because one of the things I remember just in 2R03 we use the fact that the cross

						there is a perpendicular vector. “4: A, that is true I: Sorry, you like A for what reason? 4: The thumb rule. So, it gives you a perpendicular vector.”	product spans a plane and that its perpendicular.
B	If the cross product of two nonzero vectors is a nonzero vector, then the vectors in the product span a plane.					1: I could suppose that both u and v are both the zero vector, and 0 cross 0 is just 0, and that doesn’t necessarily span a plane, and that’s not linearly independent either. 3: Oh, its E, because one of the things I remember just in 2R03 we use the fact that the cross product spans a plane and that its perpendicular. 4: Okay, when it comes to b, they could span a plane, but they also could not. But should I look at those ambiguous situations too?	2: u and v span a plane, because if not, then the cross product would be zero. 3: ….the cross product has this property that if the two vectors lie on top of each other, then you can’t get w. So, they have to be far enough apart that the two initial vectors have to be linearly independent.
C	If the cross product of two vectors is a nonzero vector, then the two vectors and their cross product are linearly independent.					3: Because w is perpendicular to u and v, u, v and w have to be linearly independent […] “4: I think I’ll go with A, B and C. I: Okay. And you’re saying C is true for what reason? 4: I think I remember Gregory Cousins mentioning this in	“I: Can you tell me why u, w and v are linearly independent? 2: Okay, well the first part I know is that u and v cannot be colinear, because if they were colinear, then I could write v as a scalar multiple of u, and then I would have the cross product is zero. Then w is perpendicular to u and v, so

						class once. Like the answer you get from a cross product is always linearly independent to the other two vectors in the cross product.” 1: I could suppose that both u and v are both the zero vector, and 0 cross 0 is just 0, and that doesn’t necessarily span a plane, and that’s not linearly independent either.	that forms a basis for R3. So, they’re linearly independent.” 3: ….the cross product has this property that if the two vectors lie on top of each other, then you can’t get w. So they have to be far enough apart that the two initial vectors have to be linearly independent, and since w is perpendicular, then all three are linearly independent.
D	A and B						
E	A, B and C.						
General Comments							
Conclusion:	Based on the information given above, this question appears to provide valid inferences to the participants understanding of span, linear independence and the cross product with the exception of responses given by participant 4. Though the participant was correct in all three of their selections, they appeared to lack the bigger picture of the relationship between the cross product and its relationship to perpendicularity, span and independence. That said, they did not arrive at any conclusion for necessarily false reasons, and as such, the question can be passed forward without further review.						

Question 16

Suppose that u, v are two vectors in $\mathbb{R}^2$, where:

$$v = ai + bj$$
$$u = bi - aj$$

Then the projection of u on to v:

a. Has length greater than v and points in the direction of v
b. Has length less than u and points in the direction of v
c. Has length greater than v and points in the direction of u
d. Has length less than u and points in the direction of u
e. Is the zero vector v.

Option	Desired Conclusion	1	2	3	4	Latent evidence for Validity of Conclusion	Manifest Evidence for Validity of Conclusion
A	The projection of two perpendicular vectors u on to v points in the direction of v and has length less than v.					4: So, if I picture it, at u, it's like a shadow of u on to v. […] You have a light source opposite the thing you want to project on to what you want to project to. Then the shadow it makes on the surface will be the projection. So, it definitely won't have a greater length. 4: It points in the direction of v for sure, because projection of any vector on to another vector is always in the direction of the vector its projected on to.	2: Has length greater than v, well that's definitely not true. Well…No, I think that could be true, if u dot v was big enough.

B	The projection of two perpendicular vectors u on to v points in the direction of v and has length less than u.				4: It points in the direction of v for sure, because projection of any vector on to another vector is always in the direction of the vector its projected on to. 4: If you shine a light source here, then you would get a vector that's shorter than itself and in the direction of v.	2: Has length less than u…yeah, I would say this one. Yeah, because its cosine theta. So, I'll put my finger on B but not final yet.
C	The projection of two perpendicular vectors u on to v points in the direction of u and has length less than v.					2: [C is not] true. That's not what projection on to v means.
D	The projection of two perpendicular vectors u on to v points in the direction of u and has length less than u.					2: [D is not] true. That's not what projection on to v means.
E	The projection of two perpendicular vectors points is the zero vector.					3: …the length has to be less than v. Which is not an option, which means I'm starting to think it's the zero vector. 4: Definitely not the zero vector, I don't see how that happens. I can't even picture that.	1: I'm looking back at E, and I think…that can't be right, can it? Because why would it be the zero vector? That doesn't make any sense. 2: If I take the dot product then I get ab-ab, which is…oh, so it's E. 3: Because the dot product here, I mean the dot product here, would be zero, because you have ab -ba, which we know is 0.

General Comments	1: So, projection is like, kind of like a shadow of a vector. 1: I know there is a formula for projection, but I don't remember what it is. 1: I would have to think about it flipped upside down, because a projection is a shadow, but the light essentially to think about it would be coming from the ground. 2: Let me try to derive what the projection formula is. […] I know it's a scalar multiple of v because it's a projection on to v. So now I just need to figure out the magnitude… 3: …the length has to be less than v. Which is not an option, which means I'm starting to think it's the zero vector. 4: I was going to write down the formula and hen calculate it, but then I remembered that it's all conceptual. But I guess I could write this down and draw it out. […] So, if I picture it, at u, it's like a shadow of u on to v. […] Yeah, that's how I picture it. You have a light source opposite the thing you want to project on to what you want to project to. Then the shadow it makes on the surface will be the projection. So, it definitely won't have a greater length.
Conclusion: REVISED + PASSED	This question as it is phrased appears to be clear to each participant. However, because the participants failed to identify that the vectors were perpendicular as they were stated, the propositions are only being indirectly assessed. Most participants tried to think about this question computationally, and while that is an entirely worthy way of reasoning through this problem, the goal of the question was to test the understanding of projection of perpendicular vectors as a special case of projection. Therefore, the question has been rephrased to state that the two vectors are orthogonal. Furthermore, the option of being in the direction of v and having length less than v has been included since most participants were not deterred by the distractors in C and D. One distractor including the phrase "In the direction of u" has been kept in an attempt to preserve testing this. The revised question is as follows: Suppose that u, v are nonzero perpendicular vectors in $\mathbb{R}^2$. Then the projection of u on to v is: a. A nonzero vector with length greater than v and points in the direction of v b. A nonzero vector with length less than v and points in the direction of v c. A nonzero vector with length less than u and points in the direction of v d. A nonzero vector with length less than u and points in the direction of u e. The zero vector Propositions assessed are revised to the following:

	A: The projection of perpendicular vectors u on to v is a nonzero vector that has length greater than v and points in the direction of v B: The projection of perpendicular vectors u on to v is a nonzero vector that has length less than v and points in the direction of v C: The projection of perpendicular vectors u on to v is a nonzero vector that has length less than u and points in the direction of v D: The projection of perpendicular vectors u on to v is a nonzero vector that has length less than u and points in the direction of u E: The projection of perpendicular vectors is the zero vector. Since these changes are minor and in line with the reasoning presented by the interview participants, this question is passed without need of further peer review.

Appendix B: Sample Peer Evaluation

Peer evaluations served as a secondary validity test method for inventory items we were unable to test in Phase I. Responses were obtained by distributing an online form to volunteers. We provided volunteers with the test item, the list of each proposition we would like to conclude from the item, and then asked them to consider whether item responses would be considered reasonable or unreasonable evidence of the conceptual link described. More specifically, volunteers were given the following set of instructions, with the number of questions varying depending on which inventory questions volunteers felt most comfortable reviewing.

This inventory has X multiple-choice questions, each with 5 options. The correct answer is highlighted for your reference.

Each question has a set of associated conclusions that we would like to make, based on a given response. Some questions will have multiple conclusions that are being assessed, and some will have multiple options assessing an overarching conclusion.

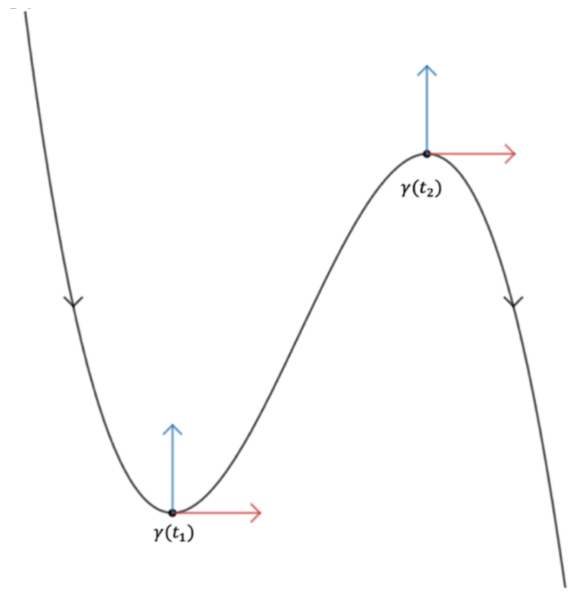
We would like you to assess whether a given response is reasonable evidence toward the desired conclusion. If you feel that the selection is not evidence of the desired conclusion, please explain why.

For each item in the form, we included a multiple-choice matrix that included each desired proposition, and the terms "unreasonable", "reasonable" and "neutral". A response for each row of the matrix, so that each proposition was assessed by the volunteer. They were also asked to explain why they felt certain options did not provide reasonable evidence for the propositions stated, where applicable, and given the opportunity to add any further comments that they felt were relevant to the test item.

This is best illustrated with an example, shown on the following page for an item included in inventory (3).

Example: Form Format and Example Response for a test item from Inventory 3

Consider the following planar curve:



Which of the following is true for the curvatures κ_1, κ_2?

a. $\kappa_1 > 0$
b. $\kappa_1 > \kappa_2$
c. κ_1 and κ_2 have the same sign
d. (a), (b) and (c)
e. (a) and (b)

Desired Conclusion(s):

By selecting (a), we would like to conclude that the student believes an anticlockwise turn of the tangent vector at a given coordinate results in positive curvature. (Correct cognitive structure)

By selecting (c), we would like to conclude that the student believes that an anticlockwise and clockwise rotation by the tangent result in the same sign of curvature, OR that the normal vector pointing in the same direction implies the same value of curvature (Incorrect cognitive structure)

By selecting (e), we would like to conclude that the student believes that and anti-clockwise rotation of the tangent vector results in positive curvature, and the clockwise rotation of the tangent vector results in a negative value of curvature. (Correct cognitive structure)

Please indicate whether you feel the selection of the options in the question provide reasonable evidence of the conclusions outlined above:

	Unreasonable	***Reasonable***	***Neutral***
Making the outlined conclusion based on the student selecting (a) would be:		**X**	
Making the outlined conclusion based on the student selecting (b) would be:		**X**	
Making the outlined conclusion based on the student selecting (c) would be:			**X**

If you answered "Unreasonable Evidence" for any of the above, please explain what aspects of the option(s) or question structure contributed to your selection.

Respondent answer: For e, the reader may have some error like replacing the curvature with e to the power of curvature, which would not be captured by the question (so that k1>k2>0). This seems unlikely though.

If you have any general comments about the question structure or options, please leave them below:

Respondent answer: [Empty]

Appendix C: Finalized Item Coverage for Inventories 2 and 3

Coverage for Inventory 2:

Question Number	Content	Question Number	Content
1	Tangent vectors at self-intersections of a parameterized curve.	**6**	Conditions for a singularity.
2	Representing the of slope for a parameterized curve.	**7**	Relationship between singularities and unit speed reparameterizations.
3	Arclength of a closed curve.	**8**	Conditions for reparameterization.
4	Relationship between the tangent vector and the normal vector at a point.	**9**	Constant curvature determines a curve in $\mathbb{R}^2$.
5	Arclength of a unit speed curve over a closed interval.		

Coverage for Inventory 3:

Question Number	Content	Question Number	Content
1	Visualizing positive and negative curvature in $\mathbb{R}^2$	**6**	Determining a curve in $\mathbb{R}^3$ with curvature and binormal conditions.
2	Relationship between signed curvature and the tangent vector.	**7**	Visualizing smooth surfaces.
3	Curvature does not determine a curve in $\mathbb{R}^3$.	**8**	Relationship between surface patches and regularity.
4	Relationship between torsion, the binormal vector, and the tangent vector.	**9**	Surface patches in $\mathbb{R}^3$.
5	Visualizing torsion in $\mathbb{R}^3$.		

Appendix D: Items in Categories (1) and (3)

Category (1)

Inventory 1 Question 7:

Consider the Riemann Sum of $x^2 - 1$ over the interval $[a, b]$:

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left(\left(-2 + \frac{4k}{n} \right)^2 - 1 \right) \times \frac{4}{n}$$

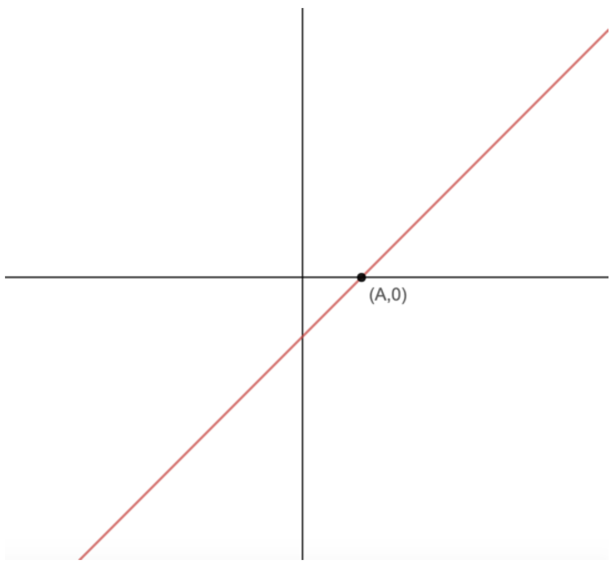
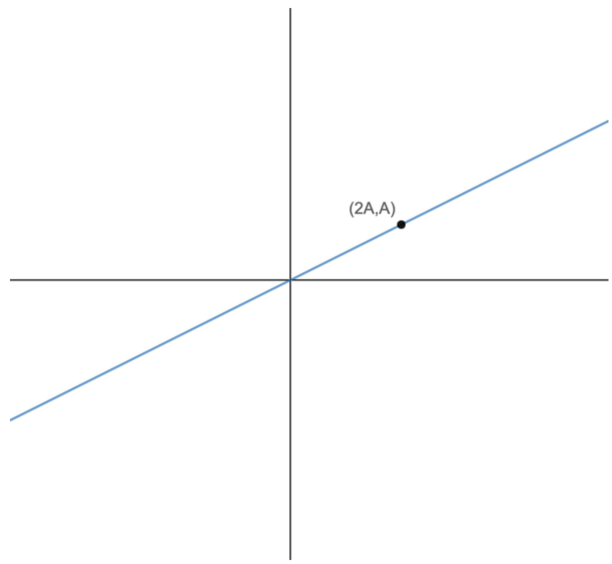
Which of the following are true statements?

a. The limit of the sum represents the signed area bounded by $x^2 - 1$ over the interval $[a, b]$.
b. The limit of the sum represents the definite integral of $x^2 - 1$ over the interval $[a, b]$.
c. The Riemann Sum is computed over the interval $[-2,2]$.
d. (b) and (c)
e. (a), (b) and (c)

	A	B	C	D	E*	Total Responses
Number of Responses	15	28	3	26	35	107

Inventory 1 Question 20:

Suppose that $f: \mathbb{R} \to \mathbb{R}$ and $g: \mathbb{R} \to \mathbb{R}$ are differentiable functions, and the graphs of their derivatives are shown below:

$f'(x)$	$g'(x)$
	

If $f(x) > 0$ and $g(x) > 0$ for all $x \in \mathbb{R}$, then which of the following are true?

a. The function $g(x)$ is increasing for all values of x
b. The function $f(x)$ has a critical point at $x = A$
c. The derivative of $g(f(x))$ is negative for $0 < x < A$
d. (b) and (c) only
e. (a), (b) and (c)

	A	B	C	D*	E	Total Responses
Number of Responses	7	28	8	40	27	110

Category (3):

Inventory 1 Question 1

Suppose $f: \mathbb{R} \to \mathbb{R}$ is a differentiable function so that $f'(x) > 0$ for all $x \in \mathbb{R}$. Which of the following must be true for the function $f(x)$?

a. $f(x)$ is onto
b. $f(x)$ is one to one
c. $f(x)$ is either one to one or onto, but not both.
d. $f(x)$ is both one to one and onto
e. Not enough information

	A	**B***	**C**	**D**	**E**	**Total Responses**
Number of Responses	15	42	4	16	32	109

Inventory 1 Question 5

Suppose that $f(x)$ is continuous over $[a, b]$. Which of the following is always true?

a. $\int_a^b f(x)dx$ is the area of the region bounded by the graph $f(x)$, the x-axis, and the lines $x = a$ and $x = b$.
b. $\int_a^b f(x)dx$ is finite.
c. $\int_a^b f(x)dx$ is an antiderivative of $f(x)$.
d. $\int_a^b f(x)dx$ may not exist.
e. None of these.

	A	**B***	**C**	**D**	**E**	**Total Responses**
Number of Responses	69	14	19	5	3	110

Inventory 1 Question 17

Suppose that the linear approximation of the function $f(x,y)\colon \mathbb{R}^2 \to \mathbb{R}$ at $(4,5)$ is given by:

$$L(x,y) = 12 + 3(x-4) + 2(y-5)$$

Which of the following is true?

a. The image of $L(x,y)$ is a plane.
b. The vectors tangent to $f(x,y)$ at $(4,5)$ are $\vec{\imath} + 3\vec{k}$ and $\vec{\jmath} + 2\vec{k}$.
c. The value of $f(5,5)$ is 15.
d. (a) and (b)
e. (a), (b) and (c)

	A	**B**	**C**	**D***	**E**	**Total Responses**
Number of Responses	21	18	7	31	32	109

Appendix E: Letter of Information, Phase I

A Study on the Position of MATH 3B03 in the Undergraduate Math Curriculum at McMaster

Student Principal Investigator:

Julie Jenkins
Department of Mathematics & Statistics
McMaster University
Hamilton, Ontario, Canada
E-mail: jenkinsj@mcmaster.ca

Faculty Supervisor:

Dr. Miroslav Lovric
Department of Mathematics & Statistics
McMaster University
Hamilton, Ontario, Canada
(905) 525-9140 ext. 27362
E-mail: lovric@mcmaster.ca

Purpose of the Study

You are invited to take part in a study whose purpose is to evaluate the conceptual framework of differential geometry. The purpose of this study is to better understand the position of MATH 3B03 in the McMaster mathematics curricula, and to assess whether alterations in instructional design contribute to enhanced student understanding.

What will happen during the study?

In this study, we will be conducting an interview in which you will be asked a series of concept-oriented questions. The questions will address concepts that have been taught in MATH 2X03 and MATH 2R03. You will be asked to describe your thought process surrounding the questions. This interview will take place via Zoom and will take approximately 60 minutes.

Zoom is an externally hosted cloud-based service. A link to their privacy policy is available here: https://zoom.us/privacy
Please note that whilst this service is approved for collecting data in this study by the McMaster Research Ethics Board, there is a small risk with any platform such as this of data that is collected on external servers failing outside the control of the research team. If you are concerned about this, we would be

happy to make alternative arrangements for you to participate, perhaps via telephone. Please talk to the researcher if you have any concerns.

Are there any risks to doing this study?

It is not likely that there will be any risks to you in this study. You might worry what we will think of you, after analyzing your surveys; or, you might be bothered by the conclusions we reach.

If you feel uncomfortable about us using your interview for research, you have an opportunity to withdraw from this research. You will be able to withdraw from the study before, during, or any time until 48 hours after your interview by sending an email to the researcher (jenkinsj@mcmaster.ca). We will destroy any data you would like excluded from the study, and never use it in our research.

When we finish our analysis and publish our findings (which we plan to conclude by 31 December 2021), we will confidentially destroy all data we have collected (the data will be destroyed in the same way as your private information, your exams, and all your work with your name and/or student ID number are destroyed on campus).

Are there any benefits to doing this study?

In this study, we hope to assess the success of courses to prepare students for the conceptual demands of upper-level mathematics courses. We would also like to identify instructional practices that benefit student understanding and suggest alternative instructional methodologies to improve the quality of student learning. It is our hope that if we can clearly exhibit the conceptual gain that interactive engagement style teaching contributes to for students. We also hope that our research, and results, will encourage faculty to consider modifying their teaching of mathematics by emphasizing student engagement in their classes. It is also possible that you may not experience any benefits from participating in the study.

Who will know what I said or did in the study?
You are participating in this study confidentially. Any data that you give us will be encoded with a unique ID for publishing. All data collected from you will be kept on a secure desktop in a password protected PDF and will be destroyed (deleted) by 31 December 2021.

Data collected throughout the course of research will be completely anonymized.

What if I change my mind about being in the study?
Your participation in this study is completely voluntary, and you will be given many opportunities to exclude yourself from the study.

If before your interview, you decide you would not like to proceed with participating, contact Julie at jenkinsj@mcmaster.ca. Your interview will be cancelled. If you decide not to be part of the study, you can contact Julie (jenkinsj@mcmaster.ca) up until 48 hours after your interview. Upon receiving notification, we will delete your data immediately.

Let us emphasize that your participation, or withdrawal from participation will not affect how you are treated in any future math course(s) that you take at McMaster.

How do I find out what was learned in this study?
The study will be complete by 1 May 2021. If you would like a brief summary of the results, please let us know during your interview when prompted.

Questions about the Study
If you have questions or need more information about the study itself, please contact Julie Jenkins at jenkinsj@mcmaster.ca. This study has been reviewed by the McMaster University Research Ethics Board and received ethics clearance.

If you have concerns or questions about your rights as a participant or about the way the study is conducted, please contact:

McMaster Research Ethics Secretariat
Telephone: (905) 525-9140 ext. 23142
c/o Research Office for Administrative Development and Support
E-mail: ethicsoffice@mcmaster.ca

CONSENT

For online studies, you will be asked if you understand following:

- I have read the information presented in the information letter about a study being conducted by Julie Jenkins of McMaster University.
- I have had the opportunity to ask questions about my involvement in this study and to receive additional details I requested.
- I understand that if I agree to participate in this study, I may withdraw from the study at any time or up until ***48 hours after my interview***
- I have been given a copy of this form.
- I agree to participate in the study.

For online studies, you will be asked to consent to the following:
1. I agree that the results from my interview will be used for research.

2. I agree that the interview can be audio recorded.

3. Would you like to receive a summary of the study results? If so, how would you like to receive it? (e.g. email, mail, etc.)

Appendix F: Oral Consent Script, Phase I

Oral Consent Script

Introduction:

Hello, I'm Julie. I am conducting interviews about the position of MATH 3B03, Geometry, in the McMaster undergraduate curriculum. I'm conducting this as part of research for a Master's Thesis at McMaster University's Department of Mathematics and Statistics in Hamilton, Ontario. I'm working under the direction Dr. Lovric of McMaster's department of Mathematics and Statistics.

Study procedures:

I'm inviting you to do a one-on-one Skype interview that will take about 60-90 minutes. I will ask you a series of multiple-choice questions about calculus and linear algebra, and to describe your problem-solving process. I will take handwritten notes to record your answers as well as use an audio recorder to make sure I don't miss what you say. We can set up a time and conferencing platform that works for us both.

Risks:
It is not likely that there will be any risks to you in this study. You might worry what we will think of you, after analyzing your surveys; or, you might be bothered by the conclusions we reach.

If you feel uncomfortable about us using your interview for research, you have an opportunity to withdraw from this research. You will be able to withdraw from the study before, during, or any time until 48 hours after your interview by sending an email to the researcher (jenkinsj@mcmaster.ca). We will destroy any data you would like excluded from the study, and never use it in our research.

Data collected throughout the course of research will be completely anonymized.

Benefits:
In this study, we hope to assess the success of courses to prepare students for the conceptual demands of upper-level mathematics courses. We would also like to identify instructional practices that benefit student understanding and suggest alternative instructional methodologies to improve the quality of student learning. It is our hope that if we can clearly exhibit the conceptual gain that interactive engagement style teaching contributes to for students. We also hope that our research, and results, will encourage faculty to consider modifying their teaching of mathematics by emphasizing student engagement in their classes.

Voluntary participation:

- Your participation in this study is voluntary.

- You can decide to stop at any time, even part-way through the interview for whatever reason, or up until 48 hours after your interview.
- If you decide to stop participating, there will be no consequences to you.
- If you decide to stop, we will destroy any data collected up to that point.
- If you do not want to answer some of the questions you do not have to, but you can still be in the study.
- If you have any questions about this study or would like more information you can call or email Julie Jenkins at (647) 546-0189 or jenkinsj@mcmaster.ca.

This study has been reviewed and cleared by the McMaster Research Ethics Board. If you have concerns or questions about your rights as a participant or about the way the study is conducted, you may contact:

McMaster Research Ethics Board Secretariat
Telephone: (905) 525-9140 ext. 23142
c/o Research Office for Administration, Development & Support (ROADS)
E-mail: ethicsoffice@mcmaster.ca

I would be pleased to send you a short summary of the study results when I finish going over our results. Please let me know if you would like a summary and what would be the best way to get this to you.

Consent questions:

- Do you have any questions or would like any additional details? *[Answer questions.]*

- Do you agree to participate in this study knowing that you can withdraw at any point with no consequences to you?
  [If yes, begin the interview.]
  [If no, thank the participant for his/her time.]

- Do you agree that the interview can be audio recorded?
  [If yes, begin the interview.]
  [If no, thank the participant for his/her time.]

- Would you like to receive a summary of the study results?
  [If yes, ask how they would like to receive the summary.]
  [If no, begin the interview.]

Appendix G: Online Survey Screening Questions, Phase I

Online Survey Screening Questions:

Instructions:
Please select either "yes" or "no" for the following questions.

Questions:

1: Have you achieved credit in MATH 2X03 or its equivalent?
Yes
No

2: Have you achieved credit in MATH 2R03, or its equivalent?
Yes
No

[If "No" for question 1 or question 2]
Thank you for being willing to participate in this study. Unfortunately, you do not meet the criteria to participate in the study at this time. If you would like to learn more about the study, please contact Julie Jenkins at jenkinsj@mcmaster.ca for more details.

[If Yes to both 1 and 2, the following two questions will be displayed]

3: Are there any other courses that are relevant to differential geometry that you have taken in the past? If so, please list them:
[Blank space for listing relevant courses]

4: Thank you for your interest in participating in this study! Please leave your email below so that the researcher (Julie) can contact you for your interview:
[Blank space for email]

Appendix H: Recruitment Email On behalf of Researcher, Phase I

Dear Students,

Julie Jenkins, a McMaster student, is asking for students to participate in a study she is doing on the conceptual framework of MATH 3B03 within the McMaster Curriculum. This research is part of her Master of Science program in Mathematics at McMaster University. Details of the study are attached, and a brief description is given below.

If you have previously achieved credit in MATH 2X03 and MATH 2R03 or equivalent, Julie is inviting you to take part in an online interview and will ask you skills and concept-based questions from previous courses you've taken. She hopes to better understand the position of MATH 3B03 in the McMaster mathematics curricula, and to assess whether variations in feedback delivery contribute to enhanced student understanding. Julie has explained that you can stop being in the study at any time during the semester. She has asked us to attach a copy of his information letter to this email. That letter gives you full details about her study.

If you are interested in getting more information about taking part in Julie's study please read the brief description (attached) and complete the online survey below. If you have questions about the study, you can contact Julie directly by using her McMaster telephone number or email address. Tel: 647-546-0189 or jenkinsj@mcmaster.ca

Your choice to participate or not will in no way change your treatment in the courses you are currently taking, or any future course you decide to take within McMaster.

Online survey
McMaster Study on MATH 3B03

In addition, this study has been reviewed and cleared by the McMaster Research Ethics Board. If you have questions or concerns about your rights as a participant or about the way the study is being conducted you may contact:

McMaster Research Ethics Board Secretariat
Telephone: (905) 525-9140 ext. 23142
Gilmour Hall – Room 305 (ROADS)
E-mail: ethicsoffice@mcmaster.ca

Sincerely,
Miroslav Lovric
Math and Stats

Appendix I: Letter of Information, Phase II

A Study on the Position of MATH 3B03 in the Undergraduate Math Curriculum at McMaster

Student Principal Investigator:

Julie Jenkins
Department of Mathematics & Statistics
McMaster University
Hamilton, Ontario, Canada
E-mail: jenkinsj@mcmaster.ca

Faculty Supervisor:

Dr. Miroslav Lovric
Department of Mathematics & Statistics
McMaster University
Hamilton, Ontario, Canada
(905) 525-9140 ext. 27362
E-mail: lovric@mcmaster.ca

Purpose of the Study

You are invited to take part in a study whose purpose is to evaluate the conceptual framework of differential geometry. The purpose of this study is to better understand the position of MATH 3B03 in the McMaster mathematics curricula, and to assess whether alterations in instructional design contribute to enhanced student understanding.

What will happen during the study?

In this study, we will be administering assignments three times during the semester. On these assignments, you will be asked a series of conceptual questions that have been addressed in MATH 2X03 and MATH 2R03, as well as some questions relating to your opinions on conceptual versus procedural understanding of mathematics. We will also monitor the progress that you make throughout the course of the semester in an attempt to isolate for the types of instruction that most benefit student learning.

These assignments are very similar to ChildsMath assignments given in other courses; you will be given a week to complete the assignment before the assignment closes. We expect that they will take approximately 30-45 minutes each, but you will be allowed as much time as necessary to complete them within the week. Please note that these assignments are mandatory for MATH 3B03, but if you elect not to have your responses used for research, you will receive full credit for completion of your assignment as usual.

Are there any risks to doing this study?

It is not likely that there will be any risks to you in this study. You might worry what we will think of you, after analyzing your assignments; or, you might be bothered by the conclusions we reach. If you feel uncomfortable about us using your assignments for research, you have an opportunity to withdraw from this research. Thus, if you initially agree to be part of the study, but then change your mind, you can withdraw at any time up until 31 December 2020 by sending an email to your course TAs, Maryam Nowroozi (nowroozm@mcmaster.ca) or Nabil Abed Allah Ali Al Asmer (alasmern@mcmaster.ca) . If you decide to withdraw from the study, you will be able to choose whether or not data that has been collected so far can continue to be used for research. Withdrawing your data from the study will have no impact on the extra credit offered by completing the assignments; that is, if you take the assignments but decide not to allow your responses to be used for research, you will still receive full credit. We will destroy any data you would like excluded from the study, and never use it in our research.

When we finish our analysis and publish our findings (which we plan to conclude by 31 December 2021), we will confidentially destroy all data we have collected (the data will be destroyed in the same way as your private information, your exams, and all your work with your name and/or student ID number are destroyed on campus).

Are there any benefits to doing this study?

In this study, we hope to assess the success of courses to prepare students for the conceptual demands of upper level mathematics courses. We would also like to identify instructional practices that benefit student understanding and suggest alternative instructional methodologies to improve the quality of student learning. It is our hope that if we can clearly exhibit the conceptual gain that interactive engagement style teaching contributes to for students. We also hope that our research, and results, will encourage faculty to consider modifying their teaching of mathematics by emphasizing student engagement in their classes. It is also possible that you will not experience any benefits as a part of this study.

Who will know what I said or did in the study?

You are participating in this study confidentially. Any data that you give us will be encoded with a unique study ID for publishing. All data collected from you will be kept on a secure desktop in a password protected PDF and will be destroyed (deleted) by 31 December 2021. Data collected throughout the course of research will be completely anonymized.

What if I change my mind about being in the study?

Your participation in this study is completely voluntary. If you decide not to be part of the study, you can contact your course TA at any time up until 31 December 2020 and indicate which data you would like to be destroyed. Upon receiving notification, we will delete your data immediately.

Nothing will happen to your assignment credit if you decide not to have your responses used for research.

How do I find out what was learned in this study?

The study will be complete by 1 May 2021. If you would like a brief summary of the results, please let us know when prompted in your sign up survey when prompted.

Questions about the Study

If you have questions or need more information about the study itself, please contact Julie Jenkins at jenkinsj@mcmaster.ca. This study has been reviewed by the McMaster University Research Ethics Board and received ethics clearance.

If you have concerns or questions about your rights as a participant or about the way the study is conducted, please contact:

McMaster Research Ethics Secretariat
Telephone: (905) 525-9140 ext. 23142
c/o Research Office for Administrative Development and Support
E-mail: ethicsoffice@mcmaster.ca

CONSENT

- I have read the information presented in the information letter about a study being conducted by Julie Jenkins of McMaster University.
- I have had the opportunity to ask questions about my involvement in this study and to receive additional details I requested.
- I understand that if I agree to participate in this study, I may withdraw from the study at any time or up until ***31 December 2020.***
    - I have been given a copy of this form.
    - I agree to participate in the study.

In your sign up survey, you will be asked the following:
Have you read and understood the Letter of Information on this study?
[ ] Yes
[ ] No

Have you read and understood the consent information on this study?
[ ] Yes
[ ] No

Would you like to participate in this study?
[ ] Yes
[ ] No

If "yes" to all three questions above in your sign up survey, you will be asked to consent to the following:

1. I agree that the researchers may obtain my mathematics course history and that it can be used for research.

[ ] Yes
[ ] No

2. I agree to have my midterm performance monitored for research.
[ ] Yes
[ ] No
4: If you would like to receive a summary of our findings, please leave your email address or mailing address below:
[Space for email address or email address.]

On your assignments, you will be given information about the study again. After this preamble information, you will then be asked:

Having read the above, I understand that by clicking the "Yes" button below, I agree to take part in this study under the terms and conditions outlined in the accompanied letter of information.

[ ] Yes, "I agree to participate"
[ ] No, "I do not agree to participate"

Note that you do not have to consent to participate for your midterms or previous academic data to be used for research to have your assignments used for research, or vice versa, and not consenting to any part of the study will not change your ability to obtain 100% credit in the course.

Appendix J: Recruitment Script for Phase II

Online Live Recruitment Script

My name is Julie Jenkins, and I am a master's student in the Department of Mathematics and Statistics. I am conducting a study on the place of MATH 3B03 in the McMaster undergraduate curriculum. The purpose of this study is to better understand the conceptual framework of MATH 3B03.

Throughout your course, you will have four online assignments worth 12% of your final grade. I am looking for students who would be willing to share their responses to those assignments and midterms with me.

The assignments are offered for credit as part of your usual coursework, but if you decide against participating you will not be penalized and will still be able to achieve 100% in the course by completing them. Indicate on your assignment that you would not like to have your data used for research, and your data will be immediately destroyed. Credit for your assignment, should you complete it, will be awarded in full regardless of whether or not you decide to participate in the study. In addition, your midterms will also be marked as usual and credit will be awarded by your TA in accordance with your performance on them, regardless of whether or not you decide to share them with me for research.

It is not likely that there will be any risks to you in this study. You might worry what we will think of you, after analyzing your surveys; or, you might be bothered by the conclusions we reach. If you feel uncomfortable about having your surveys being used for research, you can contact your course TA any time before December 31st, 2021 to withdraw from the study. Your data will be encoded with a unique study ID before I receive it; in other words, I will never know the names of any study participants, and your data will be completely anonymized. When I finish my analysis, I will confidentially destroy all data collected. I expect this will be complete by 31st of December 2021.

This study has been reviewed by the McMaster University Research Ethics Board and received ethics clearance. If you have concerns or questions about your rights as a participant or about the way the study is conducted, please contact McMaster Research Ethics Secretariat. Their office's contact information is on the letter of information that you will receive today.

If you are interested in participating, please complete the survey form that Dr. Lovric will be emailing and posting, and indicate that you consent your data to be used when prompted.

Your participation in this study is completely voluntary. If you decide not to be part of the study, you can contact your course TA at any time up until 31 December 2020 and indicate which data you would like to be destroyed. Upon receiving notification, I will delete your data immediately.

At this point, do you have any questions about the study?
[Answer questions]
Thank you for your time and consideration.

Appendix K: Recruitment Email On behalf of Researcher, Phase II

Dear Students,
Julie Jenkins, a McMaster student, is asking for students to participate in a study she is doing on the conceptual framework of MATH 3B03 within the McMaster Curriculum. This research is part of her Master of Science program in Mathematics at McMaster University. Details of the study are attached, and a brief description is given below:

Julie is inviting you to take part in a study, asking concept-based questions from previous courses you've taken and new material that you learn throughout the course. She hopes to better understand the position of MATH 3B03 in the McMaster mathematics curricula, and to assess how performance on these questions correlates to midterm achievement. Participation in the study does not require any extra action on your part, as the questions Julie is researching are embedded into this course as the online ChildsMath assignments, each of which will likely take between 30-45 min to complete. Julie has explained that you can stop being in the study at any time during the semester. She has asked us to attach a copy of her information letter to this email. That letter gives you full details about her study.

If you are interested in getting more information about taking part in Julie's study please read the brief description below and complete the online survey below by the end of the week. No further action on your part is required. If you have questions about the study, you can contact Julie directly by using her McMaster telephone number or email address. Tel: 647-546-0189 or jenkinsj@mcmaster.ca. Taking part or not taking part in this study will not affect your status in MATH 3B03, and either way you will still be able to earn 100% in MATH 3B03 without participating in the study. Your choice to participate or not will in no way change your treatment in the course, or any future course you decide to take within McMaster.

Online Sign Up Survey

Please complete the survey above by the **Friday** if you are interested in participating.

In addition, this study has been reviewed and cleared by the McMaster Research Ethics Board. If you have questions or concerns about your rights as a participant or about the way the study is being conducted you may contact:

McMaster Research Ethics Board Secretariat
Telephone: (905) 525-9140 ext. 23142
Gilmour Hall – Room 305 (ROADS)
E-mail: ethicsoffice@mcmaster.ca
Sincerely,
Dr. Miroslav Lovric
Professor, Mathematics and Statistics
McMaster University

[**Attachment: Letter of Information Phase II. See Appendix I.**]