

**What do Students Learn in the *Numbers for Life* Course at McMaster University?  
Assessing students' improvement and retention of numeracy knowledge and skills**

By  
Madison Goertz, B.Sc.

## **Abstract**

Our society is surrounded by numbers and throughout our lifetime we all experience numeric situations daily. Developing necessary numeracy skills is a crucial part of being able to fully participate in modern technological society and engage in the world around us. The course *Numbers for Life* at McMaster University (Math 2UU3) is designed to teach about critical numeracy problems that we are faced with in our daily lives and is offered to non-mathematics major students in their second year or above. Students in the course were surveyed three times through a pre-test, post-test and delayed post-test, that was written one year after course completion. Using the responses to these survey instruments, this thesis focuses on studying the retention of a student's ability to understand numeric information and their ability to communicate their answers. Having good retention is key for a learner to successfully apply what they have learned in future real-life scenarios. By studying the retention of students' responses to commonly encountered real-world math problems, we can determine how valuable courses like *Numbers for Life* are to have in place for all students.

## **Acknowledgments**

I would like to thank the Department of Mathematics and Statistics at McMaster University for funding my Master's and allowing me to continue pursuing my passion for Mathematics.

Thanks to the McMaster Athletics Department who supported me as a student-athlete, by giving me the resources to succeed on and off the field. A big shoutout to my soccer coaches, the HPA staff and my physiotherapist who continuously help me reach my goals.

Thank you to my family, who have always been there for me no matter how far away we are from each other. To my parents who give me the freedom and confidence to chase my dreams and never give up, I could not have done it without you.

From the very beginning of my academic career, I would like to thank my friends and teammates who continue to push me to be a better version of myself every day. From those who listen to me talk about math to those who inquire about my future, your support does not go unnoticed.

I am so grateful to have worked with my supervisors, Dr. Miroslav Lovric and Dr. Erin Clements, who took a chance on me and have supported me throughout this process. You both inspire me as educators and taught me to enjoy the process.

Lastly, I would like to thank Dr. Aaron Childs and Dr. Anas Abdallah for serving as my committee members for my M.Sc. thesis defense. I appreciate your attendance and the roles you had within my M.Sc. program.

## Table of Contents

<a href="#">Introduction</a> .....	4
<a href="#">Research Questions</a> .....	7
<a href="#">Literature Review</a> .....	8
<a href="#">Methodology</a> .....	14
<a href="#">Results</a> .....	18
<a href="#">ChatGPT</a> .....	29
<a href="#">Discussion</a> .....	36
<a href="#">Conclusion</a> .....	42
<a href="#">Bibliography</a> .....	44
<a href="#">Appendix A</a> .....	48
<a href="#">Appendix B</a> .....	53

## Introduction

Numbers play a vital role within all aspects of daily life whether we want them to or not. Knowing how taxes work, borrowing money, buying a car or a house, understanding how human actions affect climate change, predicting extreme weather conditions, testing for the presence of an infectious disease, etc., all involve reasoning with numbers. Our society is surrounded by systems and technologies that use numbers as a basis of their operation. For example, computer programs use a number system to perform operations, store data and communicate. Practically all jobs nowadays use computers as a tool to benefit or improve the operation of a business or a company. So the need for adults to have knowledge in understanding numbers is becoming even more essential. Every year we see technologies being continually advanced, which means that our understanding must do the same.

The ability to confidently use mathematical knowledge and skills in concrete real-world situations is one definition of numeracy [23]. Another way to think of numeracy is being mathematically literate, which is why some people use the term *quantitative literacy* interchangeably with *numeracy*. Just like literacy plays an essential role in society, so does being numerate. In so many aspects of daily life, people use numeracy skills to estimate, interpret, solve problems, provide reasoning and communicate effectively [23]. A few examples are personal finance, home building and renovation, planning a vacation, insurance documents, and understanding clauses in contracts, which all incorporate mathematical language that people need to be able to understand.

Many people assume numeracy is the same thing as mathematics but that is not true. Although they are related, they are not the same. In general, one can define mathematics as the domain of knowledge about the complexities of abstract or pure theories. On a basic level, math includes the topics of numbers, formulas, equations, spaces, etc. but it is a very broad subject, and its primary domain is that of abstraction. On the contrary, numeracy, as defined above, uses mathematical understanding to examine and critically solve problems (or tasks) in situations or contexts that we encounter in everyday life. A good way of summarizing the difference is given in Manaster (2001): “Quantitative literacy is the ability to understand and reason with numeric

information. [...] In quantitative literacy, numbers describe features of concrete situations that enhance our understanding. In mathematics, numbers are themselves the objects of study” [20].

This common misconception is part of the reason many people believe it is not necessary to be 'good' in math nor see a need to develop any mathematical skills for the future. I assume that some people see no connection of mathematics, they were taught in school or college/university, to the real world. This is where numeracy comes into play.

Numeracy has a different approach to numbers; it is the way that we think about them and understand them in real-life contexts. It is unrealistic that someone will never come across any mathematical concepts in their daily lives. A good example would be tipping at a restaurant, where you give an additional percentage of money to the server, or even budgeting for groceries every month when you calculate how much you typically spend. Everyone should be able to recognize these types of situations and be able to successfully understand their relevance and importance.

Many people claim they cannot do mathematics, let alone make sense of mathematical problems, and this will continue to be a problem unless we promote change in formal education. Everyone is capable of being numerate as well as literate, but why do we see such a drop-off when it comes to numeracy? The numeracy gap, as it is called, is the difference between necessary and actual numeracy. A survey done by the Organisation for Economic Cooperation and Development (OECD) in 2013 found that 54% of Canadian adults score below level 3 (median level to fully participate in modern technological society) for numeracy [23]. This is a major concern, given how much innovation and technological advances we see happening today. Without the necessary numeracy skills, we cannot navigate and engage in the world around us.

McMaster University offers the *Numbers for Life* course (Math 2UU3) which is open to all non-mathematics major students in their second year or above, and teaches a variety of core numeracy knowledge and skills. These include quantitative reasoning, problem-solving and critical thinking discussed in situations relating to society and daily life. The themes discussed in the course include:

- Language of mathematics and everyday language: contrasting definitions, theorems, proofs and algorithms in mathematics and real life
  - 'Quantitative' in quantitative reasoning: numbers and related concepts; elementary relationships between numbers; proportional reasoning; currency conversion and units conversions as examples of proportional reasoning; fraction, ratio and percent
    - Scientific notation; visualizing numbers in narratives (developing a feel for numbers and their size); reasoning with numbers; analyzing narratives which involve numeric information and/or logical reasoning
      - Patterns of change, in particular, linear, quadratic, cubic, and exponential
      - Models involving exponential functions, such as human population dynamics, steady growth in limited environments (demographic and economic ramifications) and exponential decay; logarithmic scale, Richter scale
      - Scatter plots and building models using linear regression; non-linear regression
      - Climate change parameters: modeling the amount of carbon dioxide in the atmosphere, surface area and thickness of Arctic ice, average global temperature
      - Discussing appropriate and inappropriate use of visual representations of information/data; interpreting information presented visually; dynamic visualization
      - Personal and general finance: managing credit card debt, loans, mortgages; financial calculators; interest, inflation, CPI (consumer price index)
      - Economy and social indicators; Gross domestic product (GDP) and gross national product/income (GNP/GNI); human development index and Gini index
      - Case studies of randomness: lottery and games of chance; probability in contexts
      - Basic probability: law of large numbers; mean, spread, box-plot diagrams; normal distribution, and reasoning based on standard deviations
      - Intuition about total probability and Bayes' theorem, applied to testing for a medical condition (false positives and false negatives); risk expressed as probability
      - Basics of statistical hypothesis testing

Learning about critical problems that we will face in the future will allow us to be confident in our ability to solve them when needed. The hope is that students who finish the course will be well-equipped to deal with certain situations and make good decisions. This

course offers content that is shown to play an essential role in managing the math-related demands of day-to-day living. One of the course's main objectives is to provide students with the tools to be constructive, critically-minded, citizens.

## Research Questions

The motivation for this thesis comes from a government-funded research project sponsored by the Higher Education Quality Council of Ontario (HEQCO). This project looks to study the development of numeracy skills among students who participate in the course *Numbers for Life* at McMaster University. The students in the course are surveyed three times through a pre-test, post-test and delayed post-test that is written one year after course completion. Through studying students' responses to questions we can draw certain conclusions about their learning and ability to apply numeracy, and, in conclusion, emphasize the importance that numeracy courses have.

The *Numbers for Life* course aims to teach about critical problems that we are faced with in our daily lives. The skills that students obtain through taking this course equip them with the ability to make reasonable decisions based on analyzing specific numeric information. This idea promotes an appreciation for (a certain type of) applied mathematics and statistics within society, as well as gives people confidence when dealing with different mathematical situations.

For my thesis, I am addressing the following research questions (the labeling of questions comes from the HEQCO research proposal):

(R1) To what extent does the course *Numbers for Life* develop students' numeracy skills?

a) How has students' ability to understand numbers (e.g., absolute and relative size, patterns) and work with numeric information (e.g., approximate, estimate, scale, visualize) changed after completing the course?

d) Are students able to better communicate their answers to quantitative reasoning problems and questions after completing the course?



(R2) To what extent are numeracy skills retained a year after the course is complete?

These questions are important when considering the significance of numeracy in our society. Studying the research questions above may give us insights into just how important these types of courses are and maybe into key areas that are still missing in the early education (i.e., primary and secondary) curriculum. Because this course is only offered in a few places in Canada, this research could promote the implementation of similar courses nationally, and possibly around the world.

## Literature Review

This thesis intends to shine light on how the course *Numbers for Life* at McMaster University develops students' numeracy skills and helps prepare them for real-life situations. By investigating students' responses to specific questions throughout the course and a period of time after completion, we may be able to prove this course provides a positive impact on its learners. Mainly focusing on a student's ability to understand numeric information and their ability to communicate their answers, I am researching whether there is high retention in those areas.

Retention is described as the ability to retain information in the mind, either short-term or long-term according to the International Journal of Evaluation and Research in Education (IJERE) [31]. The recognition of how to approach solving certain problems is possible if students can remember key ideas from their previous experiences, which is why retention is so important. It is a valuable skill for learners, enabling them to successfully apply the concepts that are taught to inside or outside classroom situations.

Problem-based learning (PBL) is often based on real-world problems as are the questions we are commonly faced with when learning to be numerate. The process of using PBL in teaching gives opportunities to students to become independent, for instance when looking to evaluate situations and present justification for their solution. Becoming and being a critical thinker is a skill that one needs to continually practice through a process of asking questions,

challenging assumptions and identifying alternatives to answers [2]. Being able to use evidence to understand, and then to communicate results, conclusions, and understandings clearly and accurately is an attribute of a critical thinker. If people continue practicing this skill it will also help build their confidence in problem solving. In general, students will not only have to remember what they have learned but they need to have the ability to transfer these concepts to new situations, ones that they did not face in their formal education [19].

The most recent study done in 2021 that uses similar methods to this thesis was done by Julius Valderama and Jubert Oligo at Nueva Vizcaya State University in the Philippines [30]. This study researched the effect of time on mathematics learning retention. Similar to what we did, there were randomized pre-test and post-tests given to the students. Specifically, this project defined knowledge retention as the difference score on the unit assessment administered after instruction (follow-up post-test) (Anderson 2007). This thesis used a post-test and delayed post-test to measure the retention of students, similar to what I looked at for Math 2UU3. The recognition of the acquired information these students retain after a year can prove to play a vital role in setting them up for future success. After all, learning retention is key for a learner to successfully apply these applications in future real-life scenarios.

One of the most common problems with retention that people encounter is motivated forgetting [15]. This idea stems from the negative experiences of learning or those who have high anxiety about a specific subject. Because mathematics is perceived as one of the most difficult subjects among students, many will have a high chance of forgetting what they learnt. Mathematics reportedly is people's most hated subject and gives many people a high degree of stress [18]. These are just a few reasons why people (unconsciously or not) force themselves to forget because they do not want to feel any of the negative emotions such as stress, anxiety, anger, etc. The implication of motivated forgetting will impact students' attitudes, learning retention and overall performance towards the subject [24]. Along with students being unmotivated to learn their most hated subject, some feel that they are just not capable.

It is a common belief that everyone has the potential to learn numeracy and if they are given the proper resources, they can even excel at it. But it takes work to practice and make

progress which is why many people give up. They might also hear (and believe) the common misconception that one is either good at math or not [12]. This statement is far from true because math is essentially its own language, and it tends to agree with the statement that “if you do not use it, then you lose it” [5]. It also does not help that calculators are an easy out of doing basic math or algebra. But if one practices the basics of numeracy in their everyday life then it will help build their confidence and allow them to have future success with numbers. If teachers also repeatedly teach the basics it allows everyone to absorb the message more clearly. Being scared of having to work with numbers should not be a concern in this day and age where numbers are a huge part of our world.

Another study that looked at students' retention of knowledge, specifically in differential equations, used post-tests and delayed post-tests one year later. They used these tests to compare the retention of conceptual and procedural knowledge, which they defined as retention of knowledge. In my research, the retention of knowledge was measured with a similar understanding. I analyzed students' responses consistently using the same rubrics focusing on the students' ability to work with numeric information and students' ability to communicate their reasoning. Thus, I looked at the conceptual understanding students showed during each test. The study done by Kwon, Rasmussen and Allen resulted in additional support that teaching conceptual understanding can lead to the retention of mathematical knowledge [17]. So with the comparison in conceptual understanding, from the post-test and delayed post-test, we can look for strong evidence that students retain the information taught to them in the *Numbers for Life* course at McMaster University.

Another study that was done by L. J. Meconi looked at different ways math can be taught at various stages and studied the students' retention [21]. After the initial problem-solving test, another test was created and administered to students four weeks later. The objective of this study was to prove that children are capable of mastering more mathematics than what was originally thought. The focus was primarily on problem-solving concepts and the results showed that students could retain them regardless of the instructional method that was used [21]. Although this paper does not focus on university-level students, its conclusion that children are more than capable of learning the necessary problem-solving skills could easily be extended to

university students. Their conclusion could be interpreted as it is not "how it is learned" but more "what is learned" that is important in creating opportunities to maintain retention within problem-solving situations. This could put even more emphasis on the necessity to have courses on numeracy at a university level.

Quantitative literacy is a fundamental skill that was defined in 1997 as five different dimensions of numeracy [6]. These include practical, use of day-to-day; civic, understanding policies; professional, skills for employment; recreational, appreciating sports/activities and cultural, part of civilization. Since then, the National Numeracy Network (NNN) has tried to continue implementing quantitative skills at lower and higher levels of education as well as among the general public [24]. Conversations need to be continually had about the importance of numeracy across all levels until there is something structured put in place. We need to recognize that there is a divide between those who are numerate and those who are not, and must make an effort to try to close this gap as soon as possible.

Comparing numerate individuals with innumerate individuals, who lack basic mathematical thinking, there are significant concerns when we consider their decision-making. A research article written by Ellen Peters discusses such comparisons. Having the numerical capability to understand situations allows you to make better judgements and decisions. In other words, numerate people make better decisions; the reason is that highly numerate individuals are more likely to retrieve and use appropriate numerical principles, making them less susceptible to framing effects [28]. Being innumerate in today's world causes numerous challenges in, for instance, financial and medical domains. It is easy to imagine the obstacles one would face under certain circumstances if one misunderstood or misinterpreted the information that was given to them. A couple of examples include splitting bar tabs, deciding which option for a mortgage to choose, understanding the results of medical testing (e.g., blood test), or even adjusting a recipe to feed a certain number of people. These situations each involve basic mathematics and mathematical reasoning, but for innumerate individuals they are challenging. The question now becomes how do we help people become more numerate so that they can make good, appropriate decisions in their lives?

There is a belief that the roots of innumeracy go back to elementary education and the way mathematics is taught. Dehaene, the author of the book "The Number Sense," suggested that children spend a large amount of time learning the mechanics of math that they may not know how to apply the concepts to adulthood [10]. This becomes an issue as well if certain life situations are never seen or practiced previously in the classroom; when they encounter these situations in real life, students often have no idea what decision to make. Incorporating applications of the basic concepts of math and giving the students time to grasp the information are both very important. Similar comments about misunderstanding the applications of math concepts are heard at various levels of education, so as educators, we need to do a better job of connecting mathematics to authentic, real-life problems.

Many studies research the best ways to teach students mathematics and numeracy. If we think about what tools we can incorporate in the classroom to increase knowledge in numeracy, the most talked about is finding interactive ways to teach students [30]. The focus on communication in mathematics is growing within classrooms. Challenging students to talk and write about their reasoning allows them to practice being mathematically literate. It also helps the students learn what they have to say when talking about the methods they use to solve a problem; it is commonly believed that the best way to learn something is to teach it to someone else. Through the explanations of procedures they employed, students become better at understanding, retaining, using and transferring their thinking [30].

The OECD recognized three components that are key in developing literacy skills: writing, discussion and reading [3]. Similarly, with quantitative literacy, these concepts need to be continually practiced. From what was previously mentioned, we discussed implementing more interactive ways that promote discussion within the classroom and the importance of writing out one's reasoning. But the component that is rarely talked about is mathematical reading. If we think about every math class we have ever taken, all of them use some sort of textbook that will be important to read and learn from. Also, most math tests involve reading questions that contain mathematical language and terminology, so the need to introduce strategies to properly read and interpret these tests is crucial. It is important to give students these strategies at an early age so that they can try them and have more practice before reaching

higher-level concepts. Within assessment situations, if students misinterpret the context of the question, they would provide answers to what they assumed the question was asking, but in reality, they were moving in the wrong direction from the start.

Outside of the classroom, having numeracy skills with the ability to write, discuss and read is linked to productivity and economic growth [31]. Many highly numerate individuals have a lower risk of poverty as they typically start higher in the labor markets. On the other hand, adults with low numeracy skills struggle to realize they have difficulties and are less likely to seek support [33]. Without the acknowledgement of any struggle, it is hard to assist those who need help.

There needs to be more mathematics-related programs or courses available that are accessible to everyone. Some of these initiatives would include numeracy workshops or courses. Many institutions have math centers, provide online resources and organize tutoring to help students. These supports are an integral part of numeracy and math remediation strategies at many colleges and universities (Dziwak, 2014). In some provinces, school budgets have been under pressure and the priority of education is often given to children rather than adults [10]. In terms of adult numeracy programs, they have been threatened in recent years through funding, but they are still very much needed.

The Conference Board of Canada found that a person should at least have level 3 numeracy to function well in Canadian society; anything lower is considered inadequate [1]. The national standards of numeracy are measured at five distinct levels; where levels 1, 2 and 3 are all an entry-level understanding of numbers, symbols and charts with the designation being made dependent on the complexity. The top levels are levels 4 and 5 which involve functional skills to select and compare relevant numerical information. In 2012 the Adult Literacy and Life Skills Survey (ALL) found that 55 percent of adults were inadequate, which was approximately a 12% increase from the previous study in 2003 [1]. Through this study, we can see that the level of numeracy in Canada is poor and also in decline within the last 20 years. It is important to realize this is an issue and take action before it gets out of hand.

As we get older and have more responsibility, our need to be able to make informed decisions directly impacts our lives. As mentioned, our confidence and ability with numbers influence the quality of our decision-making as well as impact us financially and socially. An article written by Blair and Getz states that "An individual's capacity to identify and understand quantitative situations, reason quantitatively, and communicate about the role mathematics plays in the world is essential. This quantitative literacy goes beyond basic computational skills" [6]. There is a worldwide need to be numerate and have opportunities in place for all ages so that key concepts can be continually practiced.

## **Methodology**

### Introduction

The process of studying math education looks at the social world and studies the quality of certain relationships, activities, situations or materials, which is frequently referred to as qualitative research [14]. Through this research project, I hope to gain a generalization of some understandings that will allow us to develop ways of teaching numeracy and implement certain necessary refinements.

### Goals

The goal of this thesis is to study the retention of numeracy, specifically the knowledge and skills associated, of university students enrolled in the course *Numbers for Life*. By studying the retention of students' responses to commonly encountered real-world math problems, we can determine how valuable courses like these are to have in place for everyone.

The OECD conceptualizes numeracy as "*The ability to access, use, interpret and communicate quantitative information and ideas, to engage in and manage the quantitative demands of a range of situations in adult life*" [8]. This research study will examine the short-term and long-term responses of students with this definition in mind; how well are the students in the *Numbers for Life* course able to develop the key skills that are a part of numeracy

within specific time frames? There will be specific areas I will be looking at but in general, I am looking to gauge the transferability of what is learned in the course.

Generally, knowledge is retained through building the learner's confidence in the desired subject. If a student feels they have the ability to solve a problem, they will be more likely to recall necessary information on how the problem could be solved. After the completion of the course, we hope that students will have the capacity to retain numeracy skills and confidence to be able to correctly apply them in their lives.

With these goals in mind, one of the aims of the HEQCO project is to determine how the numeracy curriculum and instructional materials developed at McMaster University can be offered for adoption by other educational institutions. Promoting these ideas could be a starting point for other institutions to create numeracy courses or activities themselves for even more students to have access to.

### Survey population- Sample

The survey population is from the students who took *Numbers for Life* in the fall of 2021. This cohort contains 384 students (out of 592) who gave consent for their pre-and post-tests to be used for research. Thus, approximately two-thirds of the class gave consent, which constitutes a fairly large sample. As I used the data from the delayed post-test, this sample was reduced to 97 students, still a large sample. (Of the 384 students, about 120 replied to our call to complete a delayed post-test (a year after the course ended), and of those, 97 gave consent for their answers to be used in this research).

### Survey Instruments

This study compares students' replies to our survey instruments and course assessments throughout three temporal points. All three surveys were administered online, with each response being recorded by the McMaster Learning Management System (Avenue). There was a pre-test, given to the students during the second week of classes in September 2021, a post-test that was



given close to the completion of the course in late November 2021, and the delayed post-test administered in December 2022, one year after the course ended.

A test-retest method was the foundation for this study, which involves administering the same test at least twice to the same group after a certain time interval has elapsed. By having multiple tests at different times it allowed us to generate a repeated experiment. This helps to keep the research controlled while staying consistent with the types of questions that were included in the tests. Also, this method looks to assure validity and reliability through a series of multiple tests. For the delayed post-test, a gift card was given out to help entice students to participate and was essential because of the importance of using one cohort for this research. Keeping the students the same throughout and only looking at one cohort guaranteed a certain degree of consistency.

The three tests contained 13 questions each, related to the course topics or relative to the situations they would have seen at some point in their lives. The questions were almost identical in the three tests; only minor modifications were made in the delayed post-test, such as changing a multiple-choice question into an "explain" question.

### Ethics Approval

Before pursuing this research, an ethics clearance from the McMaster University Research Ethics Board was obtained. Recruitment was accomplished through McMaster Ethics Board-approved means of class announcements, advertisements on the course web page, and individual emails to students. The students received the letter of information and consent form through email, as well as through an announcement in a lecture and a posting on the course web page. A graduate research assistant ensured, through standard ethics-approved routines (such as coding of names), that student's privacy is fully protected.

### Analysis

From the post-test and delayed post-test administered to students, a reliability coefficient, an index of the amount of true variance operating in a set of test scores (Aron & Aron, 2003), was calculated to indicate the relationship between the two sets of scores obtained. This was done in two ways. The first way was calculating the test-retest correlation coefficient,

represented by  $r$ , to see the relationship between scores on the post-test and delayed post-test (results in discussion). The second was calculating a Cronbach alpha for each post-test and delayed post-test, which showed whether the 8 questions within each test are correlated with each other. I compared the Cronbach alpha for each test to measure the internal consistency (results in discussion).

By studying the responses of the students who gave consent I was able to measure the development of specific competencies throughout the course using specific rubrics. I qualitatively adapted each short answer response by giving them a level/mark between 1 and 5 to see students' improvement in their knowledge of numerical data for rubric A and communication for rubric D.

The rubrics that were created to answer research questions R1 and R2 were divided into four categories:

**Category A:** A student's ability to understand numbers (e.g., relative size, absolute size, numerical patterns) and work with numeric information (e.g., approximate, scale, visualize, estimate).

**Category B:** A student's ability to use logical reasoning (such as understanding and creating cause and effect arguments) and recognize logical fallacies.

**Category C:** A student's ability to engage with multiple-step problems which require quantitative reasoning.

**Category D:** A student's ability to communicate their answers to quantitative reasoning problems and questions.

Each of these categories involves key principles that must be understood and learnt (internalized) to be numerate. For my thesis specifically, I chose to focus on questions that explore understanding numeric information, and the student's ability to reasonably communicate solutions (Categories A and D). I think that these are two areas where students struggle the most - the ability to recognize a situation and then have the ability to explain their approach and solution. Both areas take the expertise of understanding mathematical language and responding

accordingly. By analyzing students' responses within these categories we hope to gather evidence that will help us justify our viewpoint that there is a need to have courses put in place so that key numeracy concepts can be continually practiced, thus giving students appropriate numeracy skills for adulthood.

## Results

My focus is to study the retention of numeracy, specifically the knowledge and skills associated, of university students enrolled in the course *Numbers for Life*. There were eight questions from a total of thirteen questions (see Appendix A), that I analyzed in terms of the student's ability to understand numeric information and to reasonably communicate their solutions. Of these eight questions, three were multiple-choice questions.

### Q4: Multiple Choice

Referring to Appendix A, question 4 was a multiple-choice question that was formulated differently on the three tests. On the delayed post-test the question was written as:

#### Question 4

The chances of four events occurring, during a winter storm day in Ontario, are given below. Which event is the *most likely* to occur?

- 1 in 45,000 of injury in a car crash
- 10 in 400,000 of injury from slipping and falling
- 2 in 50,000 of mild to medium hypothermia
- 10 in 500,000 of injury from an exploding fireplace

On the post-test, the question was asked: "...Which event is the least likely to occur?" For every multiple choice question that was used, there was only one right answer and for this question specifically, the answer is either 10 in 500,000 or 2 in 50,000 depending on whether it asked for "least" or "most" likely. Both of these variations require students to use their understanding of ratios (or rates) to determine which probability is either the smallest or the largest.

Comparing the post-test to the delayed post-test, we found that 71 responses either remained the same or improved. This means that approximately 73% of students retained the information (and the knowledge about ratios or rates) well enough to answer the delayed post-test question correctly. When analyzing students' ability to show consistent and correct answers on both the post-test and delayed post-test, 67/97 students answered correctly on both. Looking at the pre-test responses, which asked which event was least likely to occur, 88/97 students knew the correct answer. We can assume based on this, that there was prior knowledge that these students used to answer the question correctly.

### Q3: Multiple Choice and Short Answer

Question 3 was posed as a multiple choice question in the post-test, and then changed to a short answer in the delayed post-test. In the post-test, the question was stated as:

#### Question 3

The energy demand that a person needs for its cardiovascular system to function normally is computed by multiplying their cholesterol level by 9.5 and then dividing by their heart rate. Person A and Person B have the same heart rate, but person A has lower cholesterol level. Select the correct statement.

- Person A's energy demand is larger than Person B's energy demand
- Person A's energy demand is smaller than Person B's energy demand
- Person A's energy demand is equal to Person B's energy demand

In the delayed post-test, students were asked to describe the relationship between the energy demands of Person A and Person B:

#### Question 3

The energy demand a person needs for its cardiovascular system to function normally is computed by multiplying their cholesterol level by 9.5 and then dividing by their heart rate. Person A and Person B have the same heart rate, but person A has lower cholesterol level. What is the relationship between Person A's and Person B's energy demands? Show your reasoning.

Comparing the post-test to the delayed post-test, approximately 90% of students improved or maintained the correct answer. This question focuses on the ability to understand

how the value of a fraction is affected as its numerator changes and the denominator remains fixed.

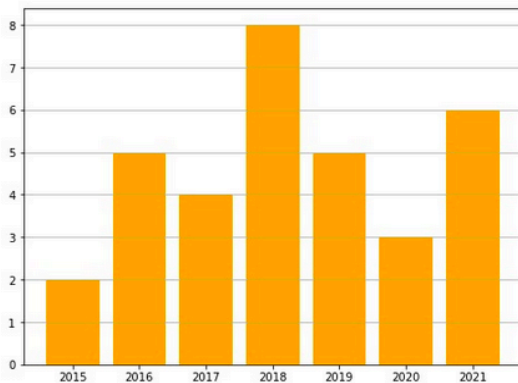
In this study, students' improvement is described as having a wrong answer on the earlier test and a correct answer on the later test. One curiosity I had was to compare the pre-test to the post-test and the pre-test to the delayed post-test to assess improvement. I found that there were no students who improved from these types of comparisons while when I compared the post-test to the delayed post-test, which was my primary focus, there were 7 students who improved. This was positive to see that after the one-year delay, several students improved their answers.

### Q5: Multiple Choice and Short Answer

Similar to question 3, the post-test of question 5 was given as a multiple-choice question, with the delayed post-test requesting an answer (either true or false) and an explanation.

#### Question 5 (post-test)

The diagram below shows the net profit (in \$ thousand) of a small family-owned company.

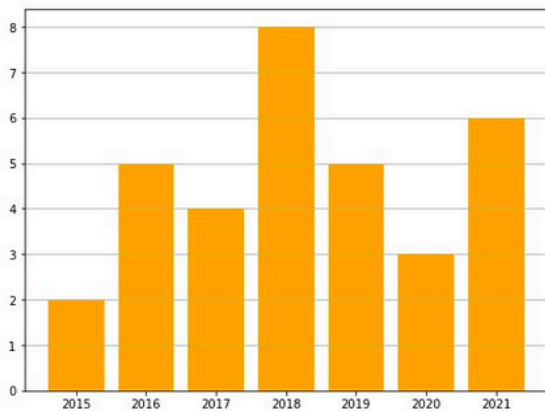


When did the company experience the largest relative growth?

- from 2015 to 2016
- from 2017 to 2018
- from 2020 to 2021

### Question 5 (delayed post-test)

The diagram below shows the net profit (in \$ thousand) of a small family-owned company.



True/False: The company experienced the largest relative growth from 2017 to 2018. State your answer as true or false, and explain your reasoning.

Correctly or not, 57 students used the word "relative" in the explanation of their solution on the delayed test, which is a desired action to use the correct language. When solely comparing the student's ability to answer correctly, from the post-test to the delayed post-test, approximately 38/97 students maintained the right answer on both tests. Those who improved from post-test to delayed post-test were 20/97 students.

This question seemed to confuse many students as there were 44 students on the post-test and 38 students on the delayed post-test who answered incorrectly. I would hypothesize that students were unsure of what the term "relative growth" means. Looking at the pre-test responses, 59 students picked the wrong answer; however, we noticed that on each subsequent test, the number of incorrect answers decreased.

The remaining questions were analyzed based on rubric A and/or rubric D (both are given in Appendix B). I compared the results of the post-test with the delayed post-test for each of these questions. After my evaluation, I asked ChatGPT (version 3.5) to perform an additional analysis using the rubrics so that I could compare the results. I wanted to determine whether this new technology can be used for this type of analysis. I share the results of the ChatGPT analysis after my initial examination.

### Q1: Using Rubrics A and D

Question 1 asks: “A sweater costs  $S$  dollars. As the price tag offers a 13% discount, you decide to buy it. At the counter, the 13% sales tax is added to the discounted price. How much will you pay for the sweater:  $S$  dollars, less than  $S$  dollars, or more than  $S$  dollars? Explain how you arrived at your answer to this question.”

Students' responses were compared using both rubrics A and D. The two components of Rubric A (A1 and A2) evaluate a student's ability to understand numbers and work with numeric information. The sum of these components was compared for both the post-test and the delayed post-test.

There were two main ways in which students attempted to solve and explain their reasoning. The most common way was by using an example and setting  $S$  to be a certain cost to see what would happen after the discount and tax. Those who answered this way were graded a maximum of 9/10 for a total of rubric A. If someone used  $S$  as a variable and showed their solution for any value of  $S$ , they received full marks, 10/10, for their answer.

A student who received full marks for each section of rubric A (A1 and A2) had the following response:

“Original price:  $S$  dollars Price with 13% discount applied:  $0.87*S$  Price with 13% sales tax applied:  $1.13*0.87*S = 0.98*S$  Therefore, you will pay less than  $S$  dollars for the sweater. Specifically, you will pay 98% of the original price” (Post-test response).

This student even concluded how much less you will pay compared to the original price. This shows both a strong ability to work with numbers (section A1) as well as a good understanding of the numeric information presented (section A2).

The average of category A on the post-test is 8.8, compared to 8.7 on the delayed post-test. This fact means that practically there is no change between the post-test and the delayed post-test; in other words, there was no drop in retention.

Rubric D evaluates students’ ability to communicate their answers to quantitative reasoning problems and questions. For category D the total average was 6.8 in the post-test compared to 7.2 in the delayed post-test. There were approximately 69% of students who scored the same or better on the delayed post-test compared to the post-test in category A and category D (viewed separately).

Comparing both categories A and D together, there was a lower percentage of improvement, with approximately 61% of students who improved or maintained their score from post-test to delayed post-test. The table below shows the averages of each category:

<b>Question 1 Averages</b>	Rubric A	Rubric A Trial 2	ChatGPT Rubric A	ChatGPT Rubric A Trial 2	Rubric D	ChatGPT Rubric D
Post-test	8.8	8.4	9.3	9.2	6.8	8.6
Delayed Post-test	8.7	7.9	8.0	9.2	7.1	8.5

**Table 1:** Question 1 average marks for the total categories in both rubrics A and D for post-test and delayed post-test. Rubric A is the sum of the average scores of the 97 responses from A1 and A2 (see Appendix B) for both the post-test and delayed post-test. Similarly, Rubric D is the sum of average scores of D1 and D2. There were two trials of marking plus two done using ChatGPT for rubric A, while rubric D had one trial using ChatGPT.

Details about the analysis using ChatGPT follow.

In question 1, students used the term “less” 93 times in the post-test and 91 times in the delayed post-test. This shows that in both tests the majority of students knew that the answer was “less than the original price.” The word “discount” was used in only 63 post-test responses, compared to the delayed post-test, which had it used 69 times. It is important, in most



mathematical responses, to explain the calculations performed; in this question, it is important to address how the discount leads us to the correct conclusion.

### Q2: Using Rubric A

The next question (question 2) asked students to calculate the number of calories consumed if they drank an entire container of milk. This question did not ask for any explanation, so I chose to only use rubric A to evaluate the responses. The question was given as follows:

#### Question 2

The table below states the nutritional facts for a container of milk.

<b>Nutrition Facts</b>	
Serving size 1 cup (220 g)	
Servings per container 3.5	
Amount per serving	
Calories 140	Calories from Fat 60
% Daily Value	
<b>Total Fat</b> 8 g	12%
Saturated fat 2g	9%
Trans fat 0 g	
<b>Cholesterol</b> 10 mg	4%
<b>Sodium</b> 235 mg	10%
<b>Total Carbohydrate</b> :	10%
Dietary Fibre 0 g	
Sugars 12 g	
<b>Protein</b> 20 g	36%

If you drink the entire container, how many calories would you consume?

By providing this specific nutrition facts chart, I thought that some students may have misunderstood the calories from the fat portion and added it to the calories per serving; however, only 2% made that mistake. Overall this question was well interpreted by students on the post-test and delayed post-test, which is good because similar labels are shown on all packaged foods in many countries.

On average, the post-test had a score of 8.9 for understanding and working with numeric information, whereas the delayed post-test had a slightly higher mean of 9.0. Approximately 75% of students maintained or improved during the one-year delay.

<b>Question 2 Averages</b>	Rubric A	ChatGPT Rubric A	ChatGPT Rubric A Trial 2
Post-test	8.9	9.0	9.0
Delayed Post-test	9.0	9.6	9.3

**Table 2:** Question 2 average levels for the total categories in rubric A for post-test and delayed post-test. Rubric A is the sum of the average scores of the 97 responses from A1 and A2 (see appendix B) for both the post-test and delayed post-test. There were two trials done using ChatGPT for rubric A (column ChatGPT Rubric A and ChatGPT Rubric A Trial 2).

When we only focus on those with the correct answer, we count 84 students on the post-test compared to 91 students on the delayed post-test.

#### Q8: Using Rubric A

Question 8 asks:

You are in New Zealand, using your Canadian phone. Assume that a roaming charge for the usage over your plan is \$6 per 80MB (megabytes) of data, and assume that you have reached the limit of your plan, meaning that you have to pay for extra data you use. You decided to watch a movie, whose size is 1.6 GB (gigabytes). How much will you pay in roaming charges for watching this movie?

This question does not ask for any justification in its solution so I only analyzed the students' ability to work with numeric information using rubric A. The average of A is 9.3 for the post-test and 9.2 for the delayed test, which is very similar so students seemed to understand how to use this numeric information. There were 81% of students who had consistently correct answers or an improvement (the same for ChatGPT).

<b>Question 8 Averages</b>	Rubric A	ChatGPT Rubric A	ChatGPT Rubric A Trial 2
Post-test	9.3	8.9	9.4
Delayed Post-test	9.2	9.0	9.0

**Table 3:** Question 8 average levels for the total categories in rubric A for post-test and delayed post-test. Rubric A is the sum of the average scores of the 97 responses from A1 and A2 (see Appendix B) for both the post-test and delayed post-test. There were two trials done using ChatGPT for rubric A (ChatGPT Rubric A and ChatGPT Rubric A Trial 2).

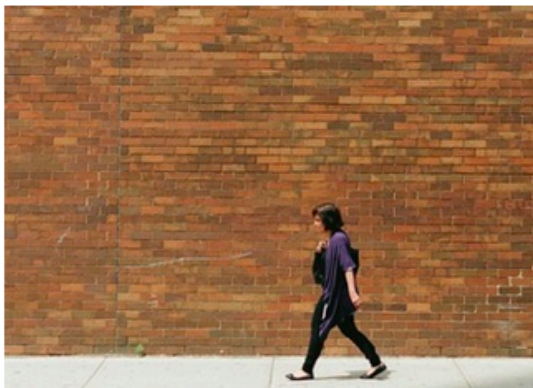
Within the students' responses to this question, the number of students who did not even mention GB (gigabytes) within their solution was 20 (post-test) vs 27 (delayed post-test) when comparing the two tests. There were 12 vs 20 of those who did not mention MB (megabytes) in their solution. To correctly answer this question, it is important to use units and units conversion.

Q9: Using Rubrics A and D

The next question is:

Question 9

Estimate the area of the part of the wall shown in this picture. (Not the area of the picture, but the area of the actual, real, wall shown below). Explain your reasoning.



The responses showed a variety of ways in which students decided to estimate the area of the wall. A few of these included using the estimated height of the woman, the brick or sidewalk dimensions, or the average stride of the woman. The majority of students chose to look up the estimates online to use in their approximation, whereas others just used their best guess.

I used rubrics A and D for the comparison of the post-test and the delayed post-test. For both tests (post/delayed), the average of category A was consistent at around 6.8. Because of the variety of solutions that could answer the question correctly, it was a lot harder to evaluate these responses, which, unfortunately, yielded a lower average. (This shows that using these rubrics (as they are) might not have been the most appropriate way to analyze this question. However, for consistency, we decided not to modify them. The major reason for this decision is that we do not compare the responses to one question with the responses to some other question, but compare the responses to the same question at two different times. In other words, the comparisons are done within the same question.) The frequency of responses that improved or remained consistently correct when analyzing the students' ability to work with numeric information was about 69%.

The averages of rubric D are 5.5 (post-test) and 5.4 (delayed post-test). With only a slight difference, many students used approximations or assumptions from their perspective. I gave higher marks to those who stated that they based their answer off of real estimates (ie., they looked up a woman's average height). As the picture may be shown at an angle, it would not be appropriate to conclude that "she looks 5 '5"; instead one could say "The average height for a woman is 162cm," and use that number in their estimation. In the post- to delayed post-comparison, 64% of students were consistent with their earlier responses or used better quality narratives in their delayed post-test solutions.

Combining both categories, A and D, our consistency (ratio between students who maintained a correct answer or improved compared to the whole sample) remains at 56% and with the abundance of ways one could answer these questions, I am not surprised with this low percentage. There were around 6 people in both tests who had either no answer or an answer that lacked any logical explanation.

Question 9 Averages	Rubric A	Rubric D
Post-test	6.8	5.5
Delayed Post-test	6.8	5.4

**Table 4:** Average levels for the total categories in both rubrics A and D for post-test and delayed post-test. Rubric A is the sum of the average scores of the 97 responses from A1 and A2 (see appendix B) for both the post-test and delayed post-test. Similarly, Rubric D is the sum of average scores of D1 and D2.

There were 80 responses in the post-test and 79 in the delayed post-test that used the woman's height to estimate the area. The second most popular way students estimated was by using dimensions of bricks; counting, it was 18 compared to 14 on each of the tests. On each of the two tests (pot/delayed), 14 students used the sidewalk to estimate.

#### Q10: Using Rubric D

The last question I looked at asked for an interpretation of the differences in the population dynamics on two islands and required an explanation rather than a numeric calculation. Thus, I chose to focus on specifically the student's ability to communicate their answers to quantitative reasoning problems and questions (rubric D). The question is:

#### Question 10

Last year, there were 100 monkeys on island X and 1000 monkeys on island Y. This year, there are 200 monkeys on island X and 1100 monkeys on island Y. Thus, on both islands, the populations of monkeys increased by 100. How would you describe what is different about the change of two populations?

There are two components of rubric D, labeled D1 and D2 (see Appendix B) that measure students' ability to communicate and use quantitative reasoning. The average level for both D1 and D2 was 9.2 in the post-test and 8.8 in the delayed post-test. In comparison, 72% of students maintained the quality of their responses or improved after the one-year delay. Looking at the table below we can see small discrepancies between ChatGPT's first trial and my scoring, with large discrepancies in the two trials performed by ChatGPT.

<b>Question 10 Averages</b>	Rubric D	ChatGPT Rubric D	ChatGPT Rubric D Trial 2
Post-test	9.2	9.1	9.5
Delayed Post-test	8.8	8.5	9.4

**Table 5:** Average levels for the total categories in rubric D for post-test and delayed post-test. Rubric D is the sum of the average scores of the 97 responses from D1 and D2 for both the post-test and delayed post-test. There were two trials done using ChatGPT for rubric D (ChatGPT Rubric D and ChatGPT Rubric D Trial 2).

Through this question, I was able to analyze students' communication by looking specifically at the language they use. A total of 39 responses used the same language on both tests to explain their solution, either with percentages or a multiplied increase. Approximately 36 students on both versions used the word "relative" in their answer in terms of relative change/growth, which is a key term in understanding this question. From both the post and delayed tests, 23 responses used the word "doubled" to describe what happened to Island X. Lastly, those responses that used "%" in their answers were 73 compared to 68.

## ChatGPT

I used ChatGPT 3.5 version to evaluate students' responses for questions 1, 2, 8 and 10 based on the rubrics A and D. There was no need for ChatGPT to evaluate the multiple choice questions (3, 4, 5) and question 9 was based on a picture in which this AI chat had no way of analyzing.

The process I followed for this analysis was sending ChatGPT the question to answer, then sending the specific part of the rubric (ie. A1) I wanted ChatGPT to use to grade. After that, I sent students' responses and asked ChatGPT to assign a grade from level 1 to level 5 based on the rubric. I will demonstrate how this was done to evaluate the responses (Question 1) for

category D1 (communication of solutions) in rubric D. Below is the conversation, with M denoting myself, giving a prompt to ChatGPT.

ChatGPT Q1: Using Rubrics A and D

M: “Answer this question: A sweater costs  $S$  dollars. As the price tag offers a 13% discount, you decide to buy it. At the counter, the 13% sales tax is added to the discounted price. How much will you pay for the sweater:  $S$  dollars, less than  $S$  dollars, or more than  $S$  dollars? Explain how you arrived at your answer to this question.”

ChatGPT: “Let's break down the problem step by step:

The sweater originally costs  $S$  dollars.

You get a 13% discount on the original price. The discounted price is then given by  $(100\% - 13\%)$  of the original price, which is 87% of  $S$ .

$$\text{Discounted Price} = 0.87 \times S$$

At the counter, a 13% sales tax is added to the discounted price. The tax is calculated as 13% of the discounted price and then added to the discounted price.

$$\text{Sales Tax} = 0.13 \times \text{Discounted Price}$$

The total amount paid is the sum of the discounted price and the sales tax:

$$\text{Total Amount} = \text{Discounted Price} + \text{Sales Tax}$$

Substituting the expressions for the discounted price and sales tax:

$$\text{Total Amount} = 0.87 \times S + 0.13 \times (0.87 \times S)$$

Now, we can simplify the expression:

$$\text{Total Amount} = 0.87 \times S + 0.1131 \times S$$

Combining the terms:

$$\text{Total Amount} = 0.9831 \times S$$

So, you will pay approximately 98.31% of the original price for the sweater after the discount and sales tax. Since this percentage is less than 100%, you will pay less than S dollars for the sweater after the discount and sales tax.”

M: “Great, now can you please analyze 97 student responses for communication of solutions to this question using levels such that:

Level 1: Student skills are emerging in this trait.

Level 2: Student can communicate using at least one correct definition of terms and concepts, or mostly correct description of terms and concepts, in their communication of solutions.

Level 3: Student can communicate using some correct definitions of terms and concepts, or mostly correct description of terms and concepts, in their communication of solutions.

Level 4: Student can communicate using mostly correct definitions of terms and concepts, or mostly correct description of terms and concepts, in their communication of solutions.

Level 5: Student can communicate using either correct definitions of terms and concepts, or correct description of meaning of terms and concepts, in their communication of solutions. I will send the responses next”

ChatGPT: “Certainly! Please provide the responses, and I'll do my best to analyze them according to the specified levels.”

M: “Please use the rubric to evaluate these responses: 1.Original price: S dollars Price with 13% discount applied:  $0.87*S$  Price with 13% sales tax applied:  $1.13*0.87*S = 0.98*S$  Therefore, you will pay less than S dollars for the sweater. Specifically, you will pay 98% of the original price.



2. less than S dollars, as the discount of 13% is first added to the original cost of the sweater. Hence, the 13% tax is added to the discounted price which makes the cost of the sweater less than the original...”

ChatGPT: “Let's evaluate each response based on the provided rubric:

Response 1: Level 5

- The student provides a clear and accurate explanation of the computational and quantitative reasoning practices. They justify their choice with reference to the context of the question and explain why their solution is reasonable.

Response 2: Level 4

- The student correctly identifies that the cost will be less than S dollars and explains the reasoning. However, the explanation lacks some detail and clarity...”

*Note: I would send approximately 20-30 responses at a time, wait for a reply and then send the next set of responses. I followed the same process for each component of the rubric (ie. A1, A2, D1, D2) for the post-test and delayed post-test responses depending on what rubric(s) I used to evaluate the question.*

For question 1 (see [Table 1](#)) there were two trials that were performed for evaluating students using rubric A. Both trials retained a total average of 9.3 on the post-test compared to 8.6 on the delayed post-test. Comparing just trial 2 when using ChatGPT, there were 73 students who had maintained or improved their solution.

The comparison based on category D using ChatGPT had a mean of 8.6 on the post-test vs 8.5 on the delayed post-test. Despite a lower average, approximately 88% of students had the same quality/correctness or better responses on the delayed test. Then combining A and D, ChatGPT saw an average of 71% of students maintaining quality/correctness or improving. Please refer to [Table 1](#) for a combined look at averages.

### ChatGPT Q2: Using Rubric A

Rubric A only was used to grade responses to question 2, because there were no explanations included in students' responses. ChatGPT generated evaluations based on the two separate components of rubric A, referred to as A1 and A2. I combined the two sums to obtain a summary evaluation for A, which can be seen in [Table 2](#). Two trials were generated for each post- and delayed post-test. Post-test had means of 9.0 in both trials, and the delayed post-test means were higher, measuring 9.6 and 9.3. The consistency from students that had a better or the same response from post-test to delayed post-test was 92%.

### ChatGPT Q8: Using Rubric A

For question 8, two trials were performed, using Rubric A. The post-test averages are 8.9 and 9.4, whereas the delayed post-test had both averages at 9.0 (see [Table 3](#)). The percentage of students' responses which remained the same or improved is 81%, which coincides with the average I found in the original analysis. After giving ChatGPT the question and rubric A component A1 (see Appendix B), here are a few of the evaluations:

M: “Please use the rubric to evaluate these responses: 1.  $80 \text{ MB} = 80 \cdot 10^6 \text{ bytes} = 8 \cdot 10^7 \text{ bytes}$   
 $1.6 \text{ GB} = 1.6 \cdot 10^9 \text{ bytes}$  Roaming Charges =  $(\$6 / 8 \cdot 10^7 \text{ bytes}) \cdot (1.6 \cdot 10^9 \text{ bytes}) = \$120$ .  
Therefore, you will pay \$120 in roaming charges for watching this movie.

2. converting GB to MB=>  $1.6 \text{ GB} = 1600 \text{ MB}$   $1600 \text{ MB} / 80 \text{ MB} = 20$  Pay in roaming charged =  $20 \cdot \$6 = \$120$

3. You will pay \$120 for watching this movie.  $1600 \text{ MB} / 80 \text{ MB} = 20$   $20 \cdot \$6 = \$120$

4. convert GB to MB:  $1.6 \text{ GB} = 1600 \text{ MB}$   $1600 / 80 = 20$   $20 \cdot \$6 = \$120$  for watching the movie

5.  $1 \text{ gb} / 1000 \text{ mb} = 1.6 \text{ gb} / x$   $x = 1600 \text{ mb}$  over-usage  $6 / 80 = \$0.075 / \text{mb}$   $0.075(1600) = \$120$  in over-use charges...”

ChatGPT: “Let's evaluate each response based on the provided rubric:

Response 1:

- Level: 5

- Comments: Correctly applies quantitative reasoning, demonstrates accurate calculations, and provides a clear explanation.

Response 2:

- Level: 4
- Comments: Demonstrates a good understanding and application of quantitative reasoning. However, the explanation could be more detailed

Response 3:

- Level: 4
- Comments: Provides a correct answer with a concise explanation. There are no major errors, but the explanation could be more detailed.

Response 4:

- Level: 4
- Comments: Correctly applies quantitative reasoning with a good explanation. One or two minor errors could be addressed.

Response 5:

- Level: 5
- Comments: Correctly applies quantitative reasoning, using a slightly different approach with the same accuracy. Provides a clear explanation...”

#### ChatGPT Q10: Using Rubric D

This question was analyzed using two trials based on rubric D; the first showed an average of 9.1 in the post-test and 8.5 in the delayed post-test, whereas the second had averages of 9.5 and 9.4 respectively. As can be seen from the averages, trial 1 and trial 2 were drastically different, with a 37% vs 86% retention rate. Looking closer, I concluded that ChatGPT's analysis of students' abilities to communicate their answers to quantitative reasoning problems was quite inadequate. I had to be more thorough when asking for ChatGPT to evaluate the students'

communication skills. For example, this is my conversation after an initial attempt of asking ChatGPT to use rubric D to evaluate responses for question 10:

M: “Please re-evaluate and give people who gave the percent increase correctly (100% or doubled vs 10%) a higher rating and if they did not mention any information on comparing both island a lower grade”

ChatGPT: “Certainly, let's reevaluate the responses with a focus on providing a higher rating for those who correctly mentioned the percent increase and clearly compared both islands. I'll adjust the ratings accordingly:

**Adjusted Ratings:**

**Responses with a higher emphasis on the percent increase and comparison:**

- Level 5: The response effectively communicates the relative change in population for both islands, using correct definitions and concepts.

**Responses without clear mention of percent increase or comparison:**

- Level 3: The response provides some correct information about the population changes but lacks a clear comparison between the two islands.

This adjusted rating focuses on giving higher grades to those who correctly mention the percent increase and make a clear comparison between both islands.”

M: “Using the rubric for evaluating students' communication of solutions can you please evaluate these responses: 1.island X's growth doubled from 100 to 200 while island Y just increased by 1.1 times.

2.The change in the two populations is different as even though both populations had an increase of 100 monkeys, the initial population of monkeys different in both populations. This means that for Island X there was a 50% increase in monkeys, however for Island Y the increase was only 9.09% as there was a large initial population of 1100 on Island Y compared to 100 monkeys on Island X...”

ChatGPT: “Let's evaluate each response based on the provided rubric:

1. Level 4: Student can communicate using mostly correct definitions of terms and concepts, or mostly correct description of terms and concepts, in their communication of solutions.
  - This response provides accurate explanations of the changes in population on both islands using appropriate percentage changes. However, it could be more concise.
2. Level 3: Student can communicate using some correct definitions of terms and concepts, or mostly correct description of terms and concepts, in their communication of solutions.
  - This response accurately describes the changes in population on both islands using percentage increases, but it lacks conciseness...”

By asking ChatGPT to evaluate based on specific language that is important to have within the solution, the results were shown to be of higher quality. Without telling ChatGPT what to give marks for, this version of AI was not able to recognize the multiple key terms that could lead to a well communicated explanation.

## **Discussion**

This thesis is part of a larger-scale project that looks to study the development of numeracy skills among students who participate in the course *Numbers for Life* at McMaster University. Previous to this study, a thesis completed by Michaela Kelly [16] used data from the pre-test and post-test to analyze students' logical thinking and ability to effectively communicate. The conclusion was that students had a good grasp of thinking and reasoning logically but lacked in the communication aspect. This was similar to my conclusion as I saw much lower average scores when scoring for communication (using rubric D) in both the post-test and delayed post-test. Michaela used interviews as a start in measuring retention of numeracy skills, however could not make significant conclusions as her sample (12 students) was too small. I took the next step, by analyzing the delayed post-test data in an attempt to arrive at a more reliable, evidence-based conclusion regarding retention of numeric reasoning and understanding as well as communication skills.

By performing my evaluation of students' responses, and asking ChatGPT to do the same, I was able to compare and determine whether ChatGPT could be a reliable means of this particular way of analyzing students' responses. Table 6 shows a summary of my and ChatGPT's evaluation results for students who either responded correctly to a question or improved their response from post-test to delayed post-test:

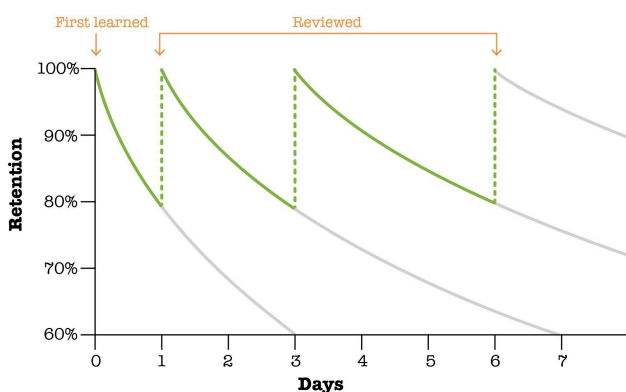
<b>Ratios of students who had correct or improved responses from post-test to delayed post-test</b>	My evaluation	ChatGPT evaluation
Question 1 Rubric A	67/97 (69.07%)	73/97 (75.26%)
Question 1 Rubric D	22/97 (22.68%)	85/97 (87.63%)
Question 2 Rubric A	73/97 (75.26%)	89/97 (91.75%)
Question 3 Multiple choice	71/97 (73.20%)	-
Question 4 Multiple choice	87/97 (89.69%)	-
Question 5 Multiple choice	58/97 (59.79%)	-
Question 8 Rubric A	79/97 (81.44%)	79/97 (81.44%)
Question 9 Rubric A	67/97 (69.07%)	-
Question 9 Rubric D	62/97 (63.92%)	-
Question 10 Rubric D	70/97 (72.16%)	83/97 (85.57%)

**Table 6:** Ratios of students who had correct or improved responses from post-test to delayed post-test compared to the total sample. Each row represents the combined totals from each rubric (ie. rubric A would be A1+A2 marks) that was used to mark each question. There was no reason to ask ChatGPT to analyze multiple-choice answers. I also did not ask ChatGPT to evaluate Question 9 because the question is based on a picture and ChatGPT has no way of analyzing pictures.

After reviewing the results for each question, I was able to determine that there is, in most cases, a high consistency that students demonstrated, either in answering both questions correctly or improving from the post-test to the delayed post-test (above 50%). These results can be summarized in [Table 6](#) for each question. Maintaining consistency in responses one year after the initial assessment is key to studying the retention of numeracy knowledge and skills. After finding a baseline of student retention, we can then implement new ways within the course to try to improve these retention rates.

The forgetting curve (see Figure 1), described by German psychologist Hermann Ebbinghaus, aims to show what happens to retention as time goes on. We can see that reviewing the learnt material multiple times leads to a slower rate of forgetting. For this study, all students come from different backgrounds and may have seen this material multiple times before, so the results will not follow this initial curve.

### Typical Forgetting Curve for Newly Learned Information



**Figure 1:** Represents the typical forgetting curve for newly learned information. This graph measures the retention of new material after a certain amount of days. The first curve is when the information is first learned and then each subsequent curve is when the material is reviewed (or otherwise visited, for instance, applied in a different context) again. The gray portion shows what is likely to happen.

When I compare my evaluation to ChatGPT's, I see a major discrepancy in question 1 rubric D (see [Table 6](#)). I found that 22/97 students either maintained a well-communicated answer or improved it from the post-test to the delayed post-test, while ChatGPT's ratio of

students was 85/97. This is the most significant difference, with my evaluation being very low compared to ChatGPT's. From my perspective, most students answered correctly but did not have good quality explanations, as outlined in rubric D. I would assume that students were not motivated to provide good quality explanations on the delayed post-test, therefore showing low improvement from post-test to delayed post-test.

But why did ChatGPT give such a high ratio? Due to this question being the first I used ChatGPT, I was not very precise about the key terms that students should have in their explanation to receive higher grades. After reflecting on the results, I can infer that ChatGPT gave high marks to students who got the correct answer without necessarily providing a well-reasoned explanation. After gaining more experience working with this natural language processing chatbot, I found a better way to ask for evaluations in communication and used that method for question 10. From Table 6 we can see that *Question 10 Rubric D* had more agreeing scores because I was more clear on how ChatGPT should evaluate students' responses. Either way, ChatGPT was not as reliable when asked to evaluate using rubric D.

While conducting this study, several limitations may have impacted the results. The participation in study was voluntary: students had to give consent for their post-test surveys to be used in this research; as well, the sample of students who took the delayed post-test was based on self-selection. Thus, there was no way for us to control the sample, for instance, to obtain a stratified sample (i.e., have subsamples of students who did not do well on the post-test, or who did well on the delayed post-test but not on the post-test, and so on). In other words, we did not know who participated in the study and whether that group well represented the entire population of all students in *Numbers for Life*. There is bias created by the people who chose to participate, and by those who did not. The participants came from different backgrounds with unique personal and learning experiences that externally impacted the results. Also, there is attrition (participant dropout) bias from post-test to delayed post-test, which causes a systematic difference between students who left and those who decided to continue and give consent.

If we think about the tests themselves, having almost identical questions on both the post-test and delayed post-test could lead to practice effects. This means that students would



perform better simply due to having seen the questions previously. This was not a serious issue for two reasons: the temporal distance between the post-test and the delayed post-test was about one year; as well, students did not know that the questions would be identical on the two tests. The responses on the pre-test showed that many students had high marks, which would be due to them having prior knowledge of certain concepts. The post-test was written at the very end of the semester (and all students had to write it), so students were likely more stressed (proximity of exams) compared to when they wrote the delayed post-test (not that close to exams). In a testing environment full of stress, students' capability to answer questions correctly decreases. Lastly, the assessment itself may not have been able to fully capture students' retention of numeracy knowledge and skills.

In research studies, instrument reliability and validity are vital to maintain. Content validity ensures test items adequately represent the content taught in the course, while reliability indicates the extent to which the results are dependable. I performed two types of analysis; one on the internal consistency of each post-test and delayed post-test, and then another to evaluate the consistency of scores over time.

Cronbach's alpha is used to measure the internal reliability of a set of items that measure the same underlying construct, concept, or idea. For this study, we have questions that measure the numeracy knowledge and skills of students in *Numbers for Life*. I calculated a Cronbach alpha for each post-test and delayed post-test, which showed to which extent the 8 questions within each test are correlated with each other. For the post-test, the calculated Cronbach alpha was 0.80 which indicates a high level of internal consistency among the questions within the test. This result also suggests that the questions within the post-test measure numeracy skills and knowledge consistently and reliably. For the delayed post-test, the Cronbach alpha was 0.55, which indicates a moderately low level of internal consistency among questions, and this lower correlation may indicate issues with reliability. Comparing both tests, we see that the post-test shows a stronger internal consistency which may indicate that the items on the post-test measure numeracy and knowledge more consistently than on the delayed post-test.

As researchers, it would be beneficial to look into the questions on the delayed post-test that could contribute to lower consistency and do revisions as needed. One way we could do this would be to perform a further item analysis or to obtain qualitative feedback from students. One limitation of Cronbach's alpha is that it assumes the items within each test measure the same core content; otherwise, Cronbach's alpha may not accurately reflect the reliability of the test.

The next method I used was the test-retest method which assesses the consistency of scores over time by administering the same test to the same group of participants at two different points in time. Because our focus is studying the long-term retention of numeracy knowledge and skills, this correlation coefficient provides insights into the relationship between scores on the post-test and delayed post-test. I found the coefficient,  $r = 0.22$ , to suggest a weak positive linear relationship between scores of the post-test and delayed post-test. They are somewhat correlated but not strongly. The positive sign of this coefficient suggests that students who performed better on the post-test tended to perform slightly better on the delayed post-test.

This method assumes that the numeracy skills being measured remain stable over the time interval, thus the longer the time interval, the lower the reliability coefficient is likely to be. Since these tests are written one year apart, there is a greater likelihood of changes in the individuals taking the test. Some factors could include additional learning, forgetting or changes in students' conditions, which may impact the correlation and affect the reliability of scores. There is no way of knowing for sure but it would still be worth examining possible modifications that could be applied to gain more consistent results.

Conducting a qualitative analysis paired with quantitative measures, and using rubrics to mark students' responses provides a comprehensive understanding of thought processes, reasoning abilities and conceptual understanding. Having a qualitative component sheds light on the learning process and insight into how students approach numeracy concepts. It was essential to incorporate this type of analysis because quantitative scores alone would not have been adequate to examine numeracy knowledge and skills.

Some of the further actions I would take if I had more time would be an analysis of predictive validity which assesses whether test scores predict future performance in relevant situations. It would be interesting to see if students who score high on the delayed post-test will perform better in numeracy-related situations in real life. Since the time frame of one year may not fully capture the retention of knowledge, doing another delayed-delayed post-test, two years after the post-test, would be valuable in understanding whether students are completely retaining the information (and to what extent the Ebbinghaus curve applies). Lastly, by performing more trials using different years of students from Math 2UU3 we would have a larger sample and be able to compare different cohorts.

## **Conclusion**

This study aimed to assess improvements in numeracy knowledge and skills by studying the students' retention in the course *Numbers for Life*. Using the post-test and delayed post-test responses we were able to analyze the consistency in correctness and the improvements of students' knowledge and skills one year after the course ended. Focusing on the question "To what extent are numeracy skills retained a year after the course is complete?" (R2), I was able to specifically look at students' ability to understand numbers, work with numeric information (Rubric A) and effectively communicate (Rubric D) which are a part of R1.

My main area of focus was measuring the retention of the numeracy skills that the *Numbers for Life* course teaches. We defined retention as the ability to retain information in the mind, either short-term or long-term, within understanding numeric information and the ability to communicate. This study of numeracy knowledge and skills for students in *Numbers for Life* revealed that the majority of students maintained their understanding of concepts one year after the course ended. Irrespective of whether this course taught students new content or served as a review of previously learned material and skills, many students were able to retain the ability to use numeric information and communicate effectively. Since good retention is key for a learner to successfully apply what they have learned in future real-life scenarios, we can deduce that the

majority of students in *Numbers for Life* will be able to do so. The retention argument can be used to support creating numeracy-type courses in other universities.

The analysis using ChatGPT version 3.5 showed similar results to my evaluation for rubric A and (sometimes serious) discrepancy when evaluating rubric D. The conclusion for future use of ChatGPT for this type of study would be to use specific language in how/what we want the natural language processing chatbot to evaluate. ChatGPT seems to be a reliable source for marking numerical ability but has a hard time judging the quality of students' communication answers. This may be an issue only for this version of ChatGPT but I do not think it does a good enough job analyzing mathematical language unless we specifically ask it to look for that language. Of course, the experience working with more advanced versions (such as ChatGPT 4.0) could have been different.

This course uses a problem-based learning approach to encourage students to assess situations and present justification for their solutions. Asking students to explain their answers and support with correct, logical reasoning and evidence makes them better at retaining and transferring their thinking [30]. We hope that through Math 2UU3 students become more confident in their ability to work with numbers and will be less hesitant when they encounter numeric situations. Building critical thinkers who can read mathematical language and understand numeric information will certainly decrease the numeracy gap.

Based on our results, *Numbers for Life* at McMaster University serves to benefit students in learning numeracy and allows many students to retain the ability to answer numeric questions one year later. Whether there are external factors contributing to students' successful retention or not, this course saw a high number of students who consistently answered correctly or improved from post-test to delayed post-test. This is a valuable course for students to gain exposure to authentic real-world applications and will be a good foundation for continuing to improve the numeracy skills of future generations.

## Bibliography

- [1] Adults With Inadequate Numeracy Skills. (n.d.). The Conference Board of Canada. <https://www.conferenceboard.ca/hcp/adlt-lownum-asp/>
- [2] Ahlam E. and Gaber, H. "Impact of problem-based learning on students critical thinking dispositions, knowledge acquisition and retention," *Journal of Education and Practice*, vol. 5, no. 14, pp. 74-83, 2014.
- [3] Beaudine, G. (2018). From Mathematical Reading to Mathematical Literacy. *Mathematics Teaching in the Middle School*, 23(6), 318–323. <https://doi.org/10.5951/mathteacmidscho.23.6.0318>
- [4] Bernhard, A. (2018). Why it matters if we become innumerate. BBC News. <https://www.bbc.com/worklife/article/20180706-why-it-matters-if-we-become-innumerate>
- [5] Bethany (2015, August 5). Math is a Foreign Language: So Treat it Like One! <https://mathgeekmama.com/math-is-a-foreign-language-so-treat-it-like-one/>
- [6] Blair R., & Getz, A. (n.d.). A Brief History of the Quantitative Literacy Movement. Carnegie Foundation for the Advancement of Teaching. Retrieved January 18, 2024, <https://www.carnegiefoundation.org/blog/a-brief-history-of-the-quantitative-literacy-movement/#>
- [7] Boylan, H. R., & Nolting, P. (2011). Improving Success in Developmental Mathematics: An Interview with Paul Nolting. In *Source: Journal of Developmental Education* (Vol. 34, Issue 3). <https://files.eric.ed.gov/fulltext/EJ986275.pdf>
- [8] Brunwell, S., & MacFarlane, A. (2020). Improving Numeracy Skills of Postsecondary Students: What is the Way Forward? Toronto: Higher Education Quality Council of Ontario. [https://heqco.ca/wp-content/uploads/2020/07/Formatted\\_Numeracy\\_FINAL.pdf](https://heqco.ca/wp-content/uploads/2020/07/Formatted_Numeracy_FINAL.pdf)
- [9] Catriona Windisch, H. (2016). How to motivate adults with low literacy and numeracy skills to engage and persist in learning: A literature review of policy interventions. *Education*, 62(3), 279–297. <https://doi.org/10.1007/sl>
- [10] Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press. <http://cognitionandculture.net/wp-content/uploads/the-number-sense-how-the-mind-creates-mathematics.pdf>

- [11] Dingwall, J. (2000). Improving Numeracy in Canada. Government of Canada Publications.  
[https://publications.gc.ca/collections/collection\\_2008/hrsdc-rhdsc/HS38-13-2000E.pdf](https://publications.gc.ca/collections/collection_2008/hrsdc-rhdsc/HS38-13-2000E.pdf)
- [12] Dunn, A. (2012). Achieving a “Number Happy” Outcome (Vol. 41, Issue 4).  
<https://www.jstor.org/stable/24767486>
- [13] Edwards, C. M., Rule, A. C., & Boody, R. M. (2017). International Forum of Educational Technology & Society Middle School Students’ Mathematics Knowledge Retention: Online or Face-to-Face Environments. Source: Journal of Educational Technology & Society, 20(4), 1–10. <https://doi.org/10.2307/26229200>
- [14] Fraenkel J., & Wallen, N. (2009). Design and Evaluate Research in Education (M. Ryan, Ed.; 7th ed.). Mejjia, Beth.
- [15] Groome, D. Eysenck, M. and Law, R. “Motivated forgetting: forgetting what we want to forget,” in M.W. Eysenck and D.H. Groome, ed., Forgetting: Explaining Memory Failure. Sage, 2020, pp. 147-168.
- [16] Kelly, M. (2023). Assessing the Improvement in Logical Reasoning of Students Enrolled in “Numbers for Life” Course at McMaster University. MSc thesis, McMaster University. Unpublished.
- [17] Kwon, O., Rasmussen, C., & Allen, K. (2005). *Students' Retention of Mathematical Knowledge and Skills in Differential Equations*. The H.W. Wilson Company. Pp. 227-239.
- [18] Ling, J. “Are Students Motivated to Forget Math?” Master Thesis, University of California, United States, 2016. [Online]. Available:  
<https://escholarship.org/uc/item/3jj1t5hp>
- [19] Lutz, S. and Huitt, W. “Information Processing and Memory: Theory and Applications,” in W. Huitt, Ed., *Becoming a Brilliant Star: Twelve core ideas supporting holistic education*. La Vergne, TN: IngramSpark, 2018, pp. 25-43.
- [20] Manaster, A. B. (2001). Mathematics and numeracy: Mutual reinforcement. In L. A. Steen (Ed.), *Mathematics and Democracy; The Case for Quantitative Literacy* (pp. 67-72). The Woodrow Wilson National Foundation Fellowship.

- [21] Meconi, L. J. (1967). Concept Learning and Retention in Mathematics. In Source: The Journal of Experimental Education (Vol. 36, Issue 1).  
<https://files.eric.ed.gov/fulltext/EJ1327374.pdf>
- [22] Nam Kwon, O., Rasmussen, C., & Allen, K. (n.d.). Students' Retention of Mathematical Knowledge and Skills in Differential Equations. <https://eric.ed.gov/?id=EJ713317>
- [23] Orpwood, G., & Sandford Brown, E. (n.d.). CLOSING THE NUMERACY GAP AN URGENT ASSIGNMENT FOR ONTARIO OCTOBER 2015.  
[http://www.numeracygap.ca/assets/img/Closing\\_the\\_numeracy\\_gap\\_V4.pdf](http://www.numeracygap.ca/assets/img/Closing_the_numeracy_gap_V4.pdf)
- [24] "Our Vision: Numeracy for all." The National Numeracy Network.  
<https://www.nnn-us.org/About-the-NNN>
- [25] Paul, R. and Elder, L. Critical thinking: Tools for taking charge of your learning and your life. Pearson, 2020.
- [26] Peng, P. Barnes, M. Namkung, J. and Sun, C. "A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics," Journal of Educational Psychology, vol. 108, no. 4, 2016, doi: 10.1037/edu0000079.
- [27] Penazzi, D. (n.d.). Retention in Mathematics students: problems and possible approaches.  
<https://pops.uclan.ac.uk/index.php/ujpr/article/view/329>
- [28] Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and Decision Making. In Source: Psychological Science (Vol. 17, Issue 5).  
<https://pubmed.ncbi.nlm.nih.gov/16683928/>
- [29] Schleicher, A. (n.d.). Why Is Numeracy Important. National Numeracy.  
<https://www.nationalnumeracy.org.uk/what-numeracy/why-numeracy-important#:~:text=Andreas%20Schleicher.OECD&text=Good%20numeracy%20is%20essential%20to,we%20need%20to%20be%20numerate.>
- [30] Thompson, D. R., & Rubenstein, R. N. (2014). Literacy in language and mathematics: More in common than you think. Journal of Adolescent and Adult Literacy, 58(2), 105–108. <https://doi.org/10.1002/jaal.338>

- [31] Valderama, J., & Oligo, J. (2021). Learning retention in mathematics over consecutive weeks: Impact of motivated forgetting. *International Journal of Evaluation and Research in Education*, 10(4), 1245–1254. <https://doi.org/10.11591/IJERE.V10I4.21577>
- [32] Why Numeracy Matters? (2001). Formal Papers of the National Forum on QL. <https://maa.org/sites/default/files/pdf/QL/WhyNumeracyMatters.pdf>
- [33] Windisch, Hendrickje Catriona. How to Motivate Adults with Low Literacy and Numeracy Skills to Engage and Persist in Learning: A Literature Review of Policy Interventions. *International Review of Education / Internationale Zeitschrift Für Erziehungswissenschaft / Revue Internationale de l'Education*, vol. 62, no. 3, 2016, pp. 279–97. JSTOR, <http://www.jstor.org/stable/24756384>



## Appendix A (Test Questions)

### QUESTIONS ON ALL PRE- POST- TESTS

1. Pre-test 2021Fall = Post-test 2021Fall

A sweater costs  $S$  dollars. As the price tag offers a 13% discount, you decide to buy it. At the counter, the 13% sales tax is added to the discounted price. How much will you pay for the sweater:  $S$  dollars, less than  $S$  dollars, or more than  $S$  dollars? Explain how you arrived at your answer to this question.

Pre-test 2022Fall = Post-test 2022Fall = Delayed Post-test Dec 2022

You are thinking of buying a sweater, and you look at the price tag. Next to it, a sign says that there is a 13% discount on all sweaters, and you decide to buy it. At the counter, the 13% sales tax is added to the discounted price. Will you pay more, less, or the same as the price seen on the price tag? Explain how you arrived at your answer.

2. Pre-test 2021Fall = Post-test 2021Fall = Pre-test 2022Fall = Post-test 2022Fall = Delayed Post-test Dec 2022

The table below states the nutritional facts for a container of milk.

Nutrition Facts	
Serving size 1 cup (220 g)	
Servings per container 3.5	
Amount per serving	
Calories 140	Calories from Fat 60
% Daily Value	
<b>Total Fat</b> 8 g	12%
Saturated fat 2g	9%
Trans fat 0 g	
<b>Cholesterol</b> 10 mg	4%
<b>Sodium</b> 235 mg	10%
<b>Total Carbohydrate</b>	10%
Dietary Fibre 0 g	
Sugars 12 g	
<b>Protein</b> 20 g	36%

If you drink the entire container, how many calories would you consume?

3. Pre-test 2022Fall = Post-test 2022Fall = Delayed Post-test Dec 2022

The energy demand a person needs for its cardiovascular system to function normally is computed by multiplying their cholesterol level by 9.5 and then dividing by their heart rate. Person A and Person B have the same heart rate, but person A has lower cholesterol level. What is the relationship between Person A's and Person B's energy demands? Show your reasoning.

Pre-test 2021Fall = Post-test 2021Fall

The energy demand that a person needs for its cardiovascular system to function normally is computed by multiplying their cholesterol level by 9.5 and then dividing by their heart rate. Person A and Person B have the same heart rate, but person A has lower cholesterol level. Select the correct statement.

- Person A's energy demand is larger than Person B's energy demand
- Person A's energy demand is smaller than Person B's energy demand
- Person A's energy demand is equal to Person B's energy demand

4. Pre-test 2021Fall = Post-test 2021Fall = Pre-test 2022Fall = Post-test 2022Fall = Delayed Post-test Dec 2022

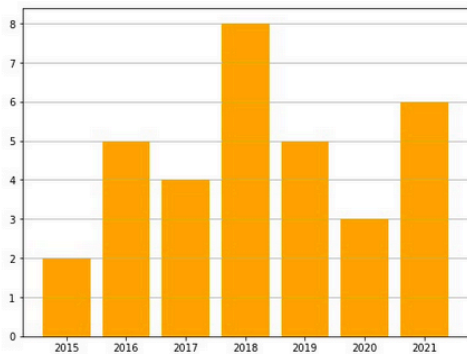
The chances of four events occurring, during a winter storm day in Ontario, are given below. Which event is the *most likely* to occur?

- 1 in 45,000 of injury in a car crash
- 10 in 400,000 of injury from slipping and falling
- 2 in 50,000 of mild to medium hypothermia
- 10 in 500,000 of injury from an exploding fireplace

[note: changed from *least likely* in 2021 to *most likely* in 2022 and delayed post-test]

5. Pretest 2022Fall = Posttest 2022Fall = Delayed Posttest Dec 2022

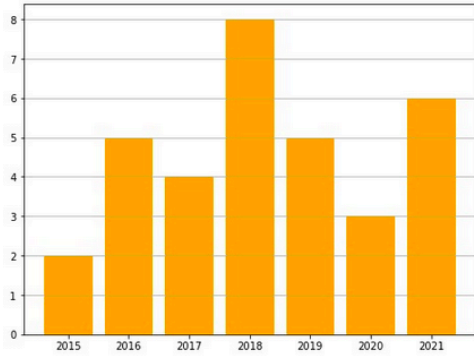
The diagram below shows the net profit (in \$ thousand) of a small family-owned company.



True/False: The company experienced the largest relative growth from 2017 to 2018. State your answer as true or false, and explain your reasoning.

Pretest 2021F = Posttest 2021F

The diagram below shows the net profit (in \$ thousand) of a small family-owned company.



When did the company experience the largest relative growth?

- from 2015 to 2016
- from 2017 to 2018
- from 2020 to 2021

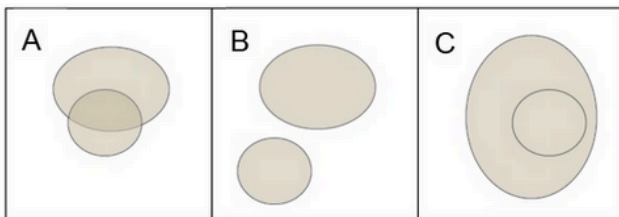
6. Pretest 2022Fall = Posttest 2022Fall = Delayed Posttest Dec 2022 = Pretest 2021F = Posttest 2021F

The sentence "The symptoms of meningitis are high fever, neck pain, and seizures" expresses a

- correlation between meningitis, and high fever, neck pain and seizures
- correlation between high fever, neck pain, and seizures
- causation, with meningitis being the cause
- causation, with high fever, neck pain, and seizures being the cause

7. Pretest 2022Fall = Posttest 2022Fall = Delayed Posttest Dec 2022

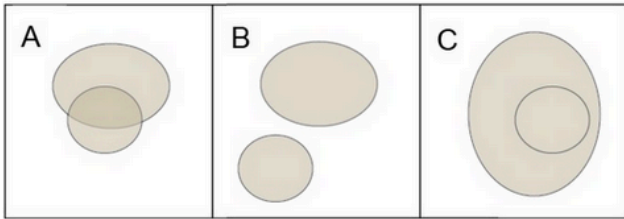
"Some animals with thick fur are mammals, and some mammals have thick fur." Which diagram represents the relationship between animals with thick fur and mammals?



Answer by stating one of A, B, or C, and justify your reasoning.

Pretest 2021F= Posttest 2021F

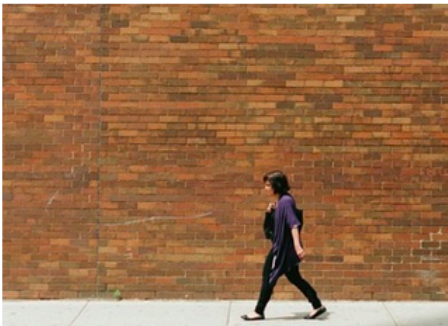
"Some animals with thick fur are mammals, and some mammals have thick fur." Which diagram represents the relationship between animals with thick fur and mammals?



- diagram A
- diagram B
- diagram C

8. Pretest 2022Fall = Posttest 2022Fall = Delayed Posttest Dec 2022 = Pretest 2021F = Posttest 2021F  
You are in New Zealand, using your Canadian phone. Assume that a roaming charge for the usage over your plan is \$6 per 80MB (megabytes) of data, and assume that you have reached the limit of your plan, meaning that you have to pay for extra data you use. You decided to watch a movie, whose size is 1.6 GB (gigabytes). How much will you pay in roaming charges for watching this movie?

9. Pretest 2022Fall = Posttest 2022Fall = Delayed Posttest Dec 2022 = Pretest 2021F = Posttest 2021F  
Estimate the area of the part of the wall shown in this picture. (Not the area of the picture, but the area of the actual, real, wall shown below). Explain your reasoning.



10. Pretest 2022Fall = Posttest 2022Fall = Delayed Posttest Dec 2022 = Pretest 2021F = Posttest 2021F  
Last year, there were 100 monkeys on island X and 1000 monkeys on island Y. This year, there are 200 monkeys on island X and 1100 monkeys on island Y. Thus, on both islands, the populations of monkeys increased by 100. How would you describe what is different about the change of two populations?

11. Pretest 2022Fall = Posttest 2022Fall = Delayed Posttest Dec 2022 = Pretest 2021F = Posttest 2021F  
Consider the statement "In my neighbourhood there are 10 dogs and they all bark at night." What would you have to do to prove that this statement is not true?

12. Pretest 2022Fall = Posttest 2022Fall = Delayed Posttest Dec 2022 = Pretest 2021F = Posttest 2021F

The statement "If you do not study, you will not pass the test" can be reinterpreted as "If you study, you will pass the test." Is this a correct conclusion? Why or why not?

13. Pretest 2022Fall= Posttest 2022Fall = Delayed Posttest Dec 2022

This ad for Harvey's burgers claims that there are *8 million ways to top your burger*.



You are suspicious of this fact, and decide to investigate. You discover that to customize your burger, you have to pick one of the three options for a bun (white, multigrain, or no bun), and then you can choose as many toppings as you wish among: 3 premium toppings, 10 other toppings and 10 sauces; for instance, you can have 2 premium toppings, 8 other toppings and 7 sauces. Based on this information you calculate the number of ways to top your burger. Does your estimate match Harvey's estimate of 8 million? Explain your reasoning.

Pretest 2021F= Posttest 2021F

Postal codes in Cook Island have the format LL-dddd where L is an uppercase letter and d is a digit; for instance, DX-3402. The first letter must be one of A, B, C, D, E, F, or G, and there are no restrictions on the second letter. The first digit in dddd cannot be 0 and cannot be 9, and there are no restrictions on the remaining three digits. What is the maximum number of postal codes available? Show how you arrived at your answer.

**Appendix B (Rubrics)**

**Rubric for Category A: Understanding and Working with Numeric Information**

**Category Description:** A student's ability to understand numbers (e.g., absolute size, relative size, patterns) and work with numeric information (e.g., approximate, estimate, scale, visualize)

	<b>Level 1</b>	<b>Level 2</b>	<b>Level 3</b>	<b>Level 4</b>	<b>Level 5</b>
<b>Ability to Understand Numbers</b>	Student skills are emerging in this trait.	Student can identify relevant numeric information but does not show interpretation in context.	Student can identify relevant numeric information from few or no distractors and shows attempt to interpret it's meaning in context.	Student can identify relevant numeric information, sometimes from other plausible distractors, and shows a valid attempt to interpret it's meaning in context.	Student can identify relevant numeric information regardless of the presence of distractors, and correctly interpret it's meaning in context.
<b>Working with Numeric Information</b>	Student skills are emerging in this trait.	Student attempts to employ an inappropriate quantitative reasoning skill but attempts to address the numeric information in the question with their answer.	Student attempts to employ an appropriate quantitative reasoning skill to work with presented numeric information in context, but makes three or more minor errors (e.g., calculation errors), or one major error (e.g., selecting an inappropriate computation).	Student attempts to employ an appropriate quantitative reasoning skill to work with presented numeric information in context, but perhaps makes one or two minor errors (e.g., calculation errors).	Student can employ quantitative reasoning skills to work with presented numeric information correctly within its context, and in a variety of forms.

## Rubric for Category D: Ability to Communicate with Quantitative Reasoning

**Category D:** A student's ability to communicate their answers to quantitative reasoning problems and questions.

	<b>Level 1</b>	<b>Level 2</b>	<b>Level 3</b>	<b>Level 4</b>	<b>Level 5</b>
<b>Communication of Solutions</b>	Student skills are emerging in this trait.	Student can communicate using at least one correct definition of terms and concepts, or mostly correct description of terms and concepts, in their communication of solutions.	Student can communicate using some correct definitions of terms and concepts, or mostly correct description of terms and concepts, in their communication of solutions.	Student can communicate using mostly correct definitions of terms and concepts, or mostly correct description of terms and concepts, in their communication of solutions.	Student can communicate using either correct definitions of terms and concepts, or correct description of meaning of terms and concepts, in their communication of solutions.
<b>Explanations of Quantitative Reasoning</b>	Student skills are emerging in this trait.	Student attempts to justify their choice of computational practices, or their quantitative reasoning practices, or why their final answer is reasonable based on the context of the question.	Student attempts to justify their choice of EITHER their computational OR quantitative reasoning practices with reference to the context of the question, and where necessary, explain why their solution is a reasonable one based on the context of the question.	Student can justify their choice of EITHER their computational OR quantitative reasoning practices with reference to the context of the question, and where necessary, explain why their solution is a reasonable one based on the context of the question.	Student can justify their choice of computational and quantitative reasoning practices with reference to the context of the question, and where necessary, explain why their solution is a reasonable one based on the context of the question.