Mathematics Word Problems: What Do Students Recall?

Tharanga Mahesh Kumara Wijetunge
Lyon College

Kirthi Premadasa
University of Wisconsin-Platteville

Mathematics instructors incorporate various word problems to illustrate the wide-ranging practical applications of mathematical concepts. Premadasa and Bhatia (2013) categorized math word problems into three main categories: I (curiosity-driven), R (relatable), and U (unrelatable). The study reveals a clear preference among students for problems in categories I and R over those in category U. Building on this categorization, we conducted a survey in which students were asked to recall the word problems they encountered throughout the semester. Our findings confirm that students notably recall and engage more with category I problems compared to those categorized as U.

Keywords: college teaching, recall ability, word problems

INTRODUCTION

To foster elevated levels of mathematical learning and student achievement, reform efforts in mathematics advocate for a range of interventions. These include active learning (Freeman et al., 2014), student-centered approaches (Lee & Hannafin, 2016), integration of technology (Premadasa et al., 2016), incorporation of real-world applications (Greer, 1997), and enhancements in teacher training (Mapolelo & Akinsola, 2015; Wijetunge et al., 2020), among others.

In this study, we investigate whether mathematics students recall specific categories of word problems more effectively by the end of a semester, focusing on the real-world applications discussed in class. Our inspiration for the current study arose from the positive effect of Realistic Mathematics Education (RME) on student enjoyment, engagement, and motivation (Aggraini & Fuzan, 2018; Tong et al., 2022; Yuanita et al., 2018). While instruction typically includes real-world applications to illustrate mathematical concepts, it's possible that students may still struggle to recall these examples when needed. Recalling real-life applications is essential for students to fully benefit from realistic mathematics education as students should be able to recognize an application context outside the classroom to fully appreciate it. If students can remember real-life applications of a certain nature better than others, teaching such applications can enhance the positive effects of RME.

Seventy-one students from two universities in Wisconsin and Arkansas participated in our study. Premadasa and Bhatia (2013) previously examined student preferences regarding word problems, noting that students tend to gravitate towards problems that provoke intrigue, curiosity, and relatability, while showing less interest in more abstract applications. In their categorization, applications that evoke intrigue and curiosity are classified under category I, those that students can relate to fall under category R, and more distant applications are grouped into category U. For instance, within the context of exponential functions, determining the age of an ancient skull would be categorized as I, while calculating the decay time of an isotope would be classified as U. Similarly, understanding the growth of a fruit fly population aligns with the category R. Authors' findings indicate that students exhibit a preference for problems categorized as I and R, such as those involving the "skull" and "fruit fly," over those classified as U, like the "isotope" problem.

The research question of this study was whether certain word problems, including those examined here or others, persist in students' short-term or long-term memory. Identifying any underlying patterns among these remembered problems can help mathematics teachers structure their applications more effectively. Teachers can enhance student learning outcomes by presenting word problems that adhere to these identified patterns when explaining mathematical concepts in the classroom.

LITERATURE SURVEY

We will start the literature survey by describing some of the benefits of RME. RME encourages students to see mathematics as a tool for solving real-world problems. According to Yuanita et al. (2018), RME's focus on practical applications and diverse representations, such as visual aids and real-life scenarios, enhances students' motivation and belief in their math abilities, fostering a positive learning environment that motivates students to engage deeply with the content.

Another study (Tong et al., 2022) examined the impact of RME on teaching complex topics, such as ellipses, to high school students in Vietnam. Findings indicate that students taught using RME not only performed better than their peers in traditional settings but also showed more positive attitudes and higher participation in math activities, underlining the model's potential to make math more engaging and accessible.

Additional studies, such as those by Anggraini and Fuzan (2018), suggest that RME helps students build confidence through active problem-solving and effective communication. By interacting with peers and teachers in practical, discussion-driven environments, students develop stronger mathematical reasoning and are more motivated to explore challenging problems, enhancing their conceptual understanding and enjoyment of mathematics.

These sources indicate that RME's emphasis on real-life context and interactive problem-solving can significantly increase student motivation, with numerous benefits for both performance and attitudes. Let us know survey some of the literature on students' recall ability.

In the 1970s, some researchers posited that students tend to forget what they learned in school shortly after final exams (Bahrick, 1979; Higbee, 1977). Implicit in this belief is the recognition that students do retain material during a semester, but the crucial inquiry is: what exactly do they retain? Is there a consistent pattern in what students remember or forget? Our study is further motivated by these questions and builds on earlier research by Premadasa and Bhatia (2013), which demonstrated that students significantly prefer word problems that intrigue (category I) or that they can easily relate to (category R) over more distant, less relatable problems (category U).

The rationale for using categories I, U, and R as the basis for our word problems stems from insights gathered in various studies. For instance, research from Ohio State University (2017) suggests that "peculiar" instances tend to be more memorable in long-term memory. This concept aligns with the "Von Restorff effect" (Von Restorff, 1933), which proposes that among a series of similar stimuli, individuals are more likely to remember those that stand out as unusual or distinctive. It is therefore valuable to explore whether these psychological findings apply within mathematics classrooms, particularly concerning word problems.

It is crucial to explore what students retain long after their school or college years, as well as what remains in their memory during a semester. Addressing these questions can offer valuable insights for students, educators, and educational institutions alike. Studies have investigated the relationship between the components of working memory and learning mathematics in both adults (Seitz & Schumann-

Hengsteler, 2002) and children (Meyer et al., 2010). Research on students' long-term and short-term memory across various subjects has been conducted (Allgood et al., 2004; Ellis & Pickard, 1977; Goldwater & Acker, 1975; VanderStoep et al., 2000). However, few studies specifically focus on mathematics.

In the following paragraphs, we delve deeper into our understanding of students' memory, particularly concerning the content they have learned. Allgood et al. (2004) surveyed graduates who had completed introductory economics courses and found varying levels of retention across different majors. Economics and Business majors reported higher levels of memorization of relevant concepts compared to other majors. Williams et al. (2005) investigated the retention of leadership competencies learned in college courses and found no statistically significant difference in retention among groups assessed one, two, or three years after completion.

In psychology, studies have reported poorer memory retention among college students for content delivered through traditional lectures (Ellis & Pickard, 1977; Goldwater & Acker, 1975). Recognition memory tests have shown better recall for extraneous material, such as jokes, compared to course content (Kintsch & Bates, 1977). In contrast, VanderStoep et al. (2000) employed free-recall techniques to investigate students' memory of current course content, finding that vivid anecdotes and demonstrations were recalled more frequently than specific course material. The relevance of the course material to the final grades was positively correlated with the frequency of recall.

Hard and Brady (2019) surveyed college seniors who had completed an introductory psychology course and found a relatively high retention rate when tested on items from their first midterm. While retention levels were higher than in previous studies, the specific types of problems or content retained were not analyzed.

These discussions highlight that students tend to recall specific instances over time, demonstrate varying levels of retention for subject-specific content, and reveal correlations between remembering course-related material and academic performance. The review also highlights the need for further research, particularly in mathematics, to identify any underlying patterns in what students retain from their coursework. Thus, we pose the question: Is there a difference in students' short-term memory of word problem categories during a semester?

METHODOLOGY

For this study, we designed a survey to assess students' retention of word problems covered in the classroom. The survey consists of two sections: a free recall section and a ranking section of a provided list of problems. Participants are students enrolled in precalculus courses at two undergraduate institutions in the USA. This study was approved by the Institutional Review Boards (IRB) of both institutions involved.

Classroom and the Content

We selected three sections of trigonometry for the study: angles and measures, right-angled trigonometry, and the law of sines. During the instruction of each section, we presented three chosen word problems to the class, one from each of the categories I, R, and U, integrated into the regular classroom activities without any unusual emphasis (see Appendix A). Additionally, we assigned six-word problems as regular homework for each section, comprising two problems from each category, alongside other non-word problems. For the unit exams, we evenly distributed problems from each category to maintain control over the study variables. This selection of problems is based on previous findings that students exhibit a preference bias for the I and R categories (Premadasa & Bhatia, 2013), and psychological research indicating that "bizarre" examples, related to the I category, are more likely to be retained in memory (Ohio State University, 2017; Von Restorff, 1933).

The Survey

The study concludes with a spot survey (Appendix A), administered at the end of the semester, to assess which word problems, if any, students recall from the relevant sections. We analyzed the survey results to draw possible conclusions. The first page of the survey asks students to write down any word problems (free recall) they remember from the class without any hints. After completing the free recall section, participants were asked to rank a list of word problems (Appendix A) based on their level of remembrance. The list contains 12 problems: 9 discussed in class (3 from each category) and 3 new ones (labeled as N). Problems 3(I), 5(R), and 8(U) in Appendix A are the new problems that were not covered in class or homework. Participants were not allowed to flip back to page 1 after proceeding to page 2. We will discuss the connections between the problems recalled in the free recall section and the rankings of the listed problems.

The Statistical Analysis

We created two mixed-effects logistic regression models (Named "The Full Model" and "The Reduced Model") for our statistical analysis. A mixed-effects logistic regression model with a binary output is a statistical model used to analyze data where the response variable is binary and has a hierarchical or clustered structure (categories: *I*, *R,* or *U*). We select our binary response variable as whether students recall at least one problem from each category (*I*, *R*, or *U*) or not. It was a binary variable, taking the value 1 if a student recalled at least one problem in the category and 0 otherwise. The distinctive feature of a mixed-effects model is the inclusion of random effects. Random effects capture unexplained variability at different levels of grouping in the data. In both models, we use "teacher" and "student" variables as random effect variables and the category variable as the fixed effect variable. The Full Model includes the category variable, and the Reduced Model excludes it. We then used the likelihood ratio test to determine whether students' recall of word problems is associated with the problem category (categories: *I*, *R,* or *U*).

RESULTS AND DISCUSSION

Findings of Free Recalled Problems

We surveyed ($N =$) 71 students, 39 from one institution and 32 from the other, and tabulated the results based on the categories of the problems they recalled. In the free recall portion of the survey, 49 students (23 from the first institution and 26 from the second) recalled at least one problem, while 22 students (16 from the first institution and 6 from the second) did not recall any. Overall, students recalled a total of 123 problems, 83 of which were relevant to the study. Among these relevant problems, 40 belonged to the I category, 25 to the R category, and 18 to the U category. Table 1 below summarizes the data. Compared to the findings in the literature, this study reported a higher percentage of participants recalling relevant problems.

TABLE 1
SUMMARY OF RECALLED PROBLEMS AND RELEVANCE

Institution	Total recalled problems	Relevant problems	I	R	U
		N %			
1	42	30 71	13	14	3
2	81	53 65	27	11	15

As described in the statistical analysis section, we developed two mixed-effects logistic regression models. The problem category was included as a variable in the full model but not in the reduced model. We compared the two models using the likelihood ratio test to determine if recalling word problems is associated with the problem category. The comparison yielded a likelihood ratio test statistic (LRT) of 8.44 with 2 degrees of freedom and a corresponding p-value of 0.015. This smaller p-value suggests that students' recall of word problems is indeed associated with the problem category (see Table 2 for the test statistics). This result prompted us to investigate whether there is a difference in recall rates among the problem categories. Specifically, we questioned whether students recalled problems from one category more frequently than those from other categories. To answer this, we conducted a pairwise comparison among the three problem categories: I, U, and R.

TABLE 2
LIKELIHOOD RATIO TEST STATISTICS

Item	Value
Log Likelihood for Reduced Model	-128.38
Log Likelihood for Full Model	-124.17
$LRT = -2[\ln(L_{Reduced\_Model}) - \ln(L_{Full\_Model})]$	8.42
Degree of freedom	2
p-value	0.015

The results indicated a statistically significant difference (p-value < 0.05) between the I and U categories. For instance, students tend to recall more intriguing word problems (e.g., finding the height of the Willis Tower) than unrelated problems (e.g., finding the distance of a ship from shore). This finding aligns with the literature review. However, there is no statistically significant difference between the I and R categories or the U and R categories. It appears that relatable problems (e.g., finding the angle of elevation when shooting free throws in basketball) and intriguing problems are remembered more frequently due to their familiarity and engaging nature. The lack of significant differences between the U and R categories suggests that further investigation is needed to understand this phenomenon. Table 3 summarizes the statistics of the pairwise comparisons.

TABLE 3
RESULTS OF THE PAIRWISE COMPARISON

Pairwise Comparison	Z value	p-value
$R - U$	0.642	0.797
$I - U$	2.693	0.019
$R - I$	2.125	0.085

Findings of the Ranking of the Recognition Level

In addition to the free recall problems, we asked participants to rank a given set of problems based on their level of recognition (see Page 2 of the survey in Appendix A). We did not conduct a statistical test to determine if the ranking pattern mirrored the recall pattern from Page 1 due to the difficulty of comparing numerical ratings. For example, if a student ranked three *U* category problems as 2, 2, and 2 (somewhat remember) and three *I* category problems as 4, 1, and 1 (do not remember and strongly remember), both categories would yield the same numerical sum. However, the *I* category would have two strongly remembered problems, while the *U* category would have none. Therefore, instead of grouping problems into categories, we compared each problem individually. Table 4 summarizes the corresponding data.

Based on the average rankings, the top four problems are from the *I* and *R* categories, while the bottom three are new problems not discussed in class. This indicates that students can identify problems not covered in class when presented with a list. However, it is important to note that during free recall, participants listed problems that were not discussed in class. This difference between free recall and ranking is intriguing and warrants further investigation.

The top four most-remembered problems are the height of the Willis Tower (*I*), the angle of a free throw in basketball (*R*), the distance to the moon (*I*), and the area of a little league field (*R*). The two remaining problems from these categories that did not rank as highly are designing an awning (*R*) and keeping up with the sun (*I*). This pattern, where certain problems within each category are remembered more frequently than others, aligns with the free recall analysis presented on Page 1. It suggests a deeper relationship between the nature of the problems and students' recall ability, inviting further investigation into why certain problems are more memorable than others.

TABLE 4
RANKING OF LEVEL OF REMEMBER OF THE PROBLEM

Problem (Category)	Rank 1	Rank 2	Rank 3	Rank 4	Average Rank
1(*R*)	12	35	18	6	2.25
2 (*I*)	28	26	8	9	1.97
3 (N)	0	9	21	41	3.45
4 (*I*)	21	20	20	9	2.21
5 (N)	4	3	16	48	3.52
6 (*U*)	13	24	24	10	2.44
7 (*U*)	21	19	11	18	2.31
8 (N)	4	15	30	22	2.99
9 (*R*)	12	22	12	24	2.65
10 (*I*)	4	20	23	24	2.94
11 (*R*)	23	23	13	11	2.17
12 (*U*)	11	15	23	22	2.79

CONCLUSIONS

The analysis of the results from this study aligns with existing psychological research, which suggests that people tend to remember the more "bizarre" events from a set of events over the long term. Instructors can leverage these findings to help students recall potential applications of mathematical concepts not only throughout the course but also beyond. Even if students do not recall the specifics of mathematical calculations, remembering the context can help them understand that similar mathematics can be applied to solve analogous problems.

For example, if a student remembers that trigonometry can be used to find the height of the Willis Tower, they can apply similar techniques to determine the height of a tall tree or a mountain. This can deepen students' appreciation of mathematical concepts and open up new possibilities for real-life applications.

We recommend that instructors consider incorporating more problems from the I category when discussing applications after teaching a topic. Modern textbooks already contain a diverse range of word problems at the end of each chapter, so selecting problems that capture students' interest should not be overly challenging. Although this study focused on mathematics, we believe similar patterns would likely apply to other disciplines as well.

FUTURE WORK AND LIMITATIONS

Several observations emerged during our investigation of the answers to the main research question, indicating a need for further research to fully comprehend the nature and effects of the problem categories. First, a qualitative study could provide deeper insights into the relationships between the categories *U* and *R*. Second, a similar approach could shed light on potential reasons why participants can promptly recognize new material from a given list yet recall unrelated problems during free recall.

The categorization of problems into *I*, *R*, or *U* is inherently subjective. However, we endeavored to minimize subjectivity by establishing guidelines for categorization and having two instructors independently classify and review the categorization of problems.

REFERENCES

Allgood, S., Bosshardt, W., Van Der Klaauw, W., & Watts, M. (2004). What students remember and say about college economics years later. *American Economic Review*, *94*(2), 259–265.

Anggraini, R.S., & Fauzan, A. (2020). The effect of realistic mathematics education approach on mathematical problem solving ability. *Edumatika: Jurnal Riset Pendidikan Matematika*, *3*(2), 94–102.

Bahrick, H.P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, *108*(3), 296.

Ellis, N.R., & Rickard, H.C. (1977). Evaluating the teaching of introductory psychology. *Teaching of Psychology, 4*, 128–132.

Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., & Wenderoth, M.P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415.

Goldwater, B.C., & Acker, L.E. (1975). Instructor-paced, mass-testing for mastery performance in an introductory psychology course. *Teaching of Psychology*, *2*, 152–155.

Greer, B. (1997). Modelling reality in mathematics classrooms: The case of word problems. *Learning and Instruction*, *7*(4), 293–307.

Hard, B.M., Lovett, J.M., & Brady, S.T. (2019). What do students remember about introductory psychology, years later?. *Scholarship of Teaching and Learning in Psychology*, *5*(1), 61.

Higbee, K.L. (1977). Your memory: How it works and how to improve it. Englewood Cliffs, NJ:Prentice-Hall.

Kintsch, W., & Bates, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 150–159.

Lee, E., & Hannafin, M.J. (2016). A design framework for enhancing engagement in student-centered learning: Own it, learn it, and share it. *Educational Technology Research and Development*, *64*, 707–734.

Mapolelo, D.C., & Akinsola, M.K. (2015). Preparation of mathematics teachers: Lessons from review of literature on teachers' knowledge, beliefs, and teacher education. *International Journal of Educational Studies*, *2*(1), 01–12.

Meyer, M.L., Salimpoor, V.N., Wu, S.S., Geary, D.C., & Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learning and individual differences*, *20*(2), 101–109.

Ohio State University. (2017, June 19). Why the 'peculiar' stands out in our memory. *ScienceDaily*. Retrieved from www.sciencedaily.com/releases/2017/06/170619092713.htm

Premadasa, K., & Bhatia, K. (2013). Real Life Applications in Mathematics: What Do Students Prefer? *International Journal for the Scholarship of Teaching and Learning*, *7*(2), n2. https://doi.org/10.20429/ijsotl.2013.070220

Premadasa, K., Wijetunge, T., & Bhatia, K. (2016). Using cellphones as virtual clickers in a mathematics classroom. *Electronic Journal of Mathematics & Technology*, *10*(3).

Seitz, K., & Schumann-Hengsteler, R. (2002). Phonological loop and central executive processes in mental addition and multiplication. *Psychological Test and Assessment Modeling*, *44*(2), 275–302.

Tong, D.H., Nguyen, T.T., Uyen, B.P., Ngan, L.K., Khanh, L.T., & Tinh, P.T. (2022). Realistic mathematics education's effect on students' performance and attitudes: A case of ellipse topics learning. *European Journal of Educational Research*, *11*(1), 403–421.

VanderStoep, S.W., Fagerlin, A., & Feenstra, J.S. (2000). What do students remember from introductory psychology? *Teaching of Psychology*, *27*(2), 89–92.

Von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. [The effects of field formation in the trace field]. *Psychologische Forschung [Psychological Research]* (in German), *18*(1), 299–342. doi: 10.1007/BF02409636.

Wijetunge, T.M., Spann, A.G., Warahena-Liyanage, G., & St John, D. (2020). Significance of the Length of Instructions with Technology: What Pre-Service Teachers Say? *i-Manager's Journal of Educational Technology*, *17*(1), 22–34.
Williams, J.R., Townsend, C.D., & Linder, J.R. (2005). Teaching leadership: Do students remember and utilize the concepts we teach. *Journal of Leadership Education*, *4*(1), 62–74.
Yuanita, P., Zulnaidi, H., & Zakaria, E. (2018). The effectiveness of realistic mathematics education approach: The role of mathematical representation as mediator between mathematical belief and problem solving. *PLoS One*, *13*(9), 1–20. e0204847.

APPENDIX

Word Problem Recall Survey

In the survey below, you will have questions based on the topics we did in the class which had word problems. Please answer all the questions in the survey. For the first survey questions, if you don't remember any, please feel free to say so.

List all the word problems we discussed in class throughout the semester that you can remember. If possible, identify the related mathematics concept. If you don't remember any, please feel free to say so.

……
……
……
……
……
……
……
……
……
……
……

Please read the listed word problems in the first column. Rank the word problems by how well you remember the problem's context (For example, put 1 in the box in front of the problem whose context you remember most, put 2 in front of the one you remember next, and so on in the second column.

Word Problem	**Level of Remembrance** 1 – Strongly remember 2 – somewhat remember 3 – Vaguely remember 4 – Not remember	**If possible, write something about how to solve the problem**
1) Designing a little league field (finding the area of the warning track)		
2) How tall is the Willis Tower?		
3) The diameter of a Ferris Wheel.		
4) Calculating the distance to the moon		
5) Finding the maximum profit.		
6) Using latitude information, find the distance between Memphis and New Orleans		
7) Finding the distance of a ship from shore.		
8) Approximate the area of a lake		
9) Designing an awning to account for the sun's angle		
10) How fast would you have to travel on the surface of Earth at the equator to keep up with the sun?		
11) Find the angle of elevation when shooting free throws in Basketball.		
12) Finding the height of an airplane		