Canadian Science Publishing | Canadian Journal of Chemistry

Investigation of the Correlation Between College Students' Success with Stoichiometry Subproblems and Metacognitive Awareness

Journal:	*Canadian Journal of Chemistry*
Manuscript ID	cjc-2019-0384.R3
Manuscript Type:	Article
Date Submitted by the Author:	23-Jun-2020
Complete List of Authors:	Gulacar, Ozcan; University of California, Davis, Chemistry Cox, Charlie ; Stanford University, Chemistry Tribble, Emma; University of California, Davis, Chemistry Cohen-Sandler, Rose; University of California, Davis, Chemistry Rothbart, Nicholas; University of California, Davis, Chemistry
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)
Keyword:	

Investigation of the Correlation Between College Students' Success with Stoichiometry Subproblems and Metacognitive Awareness

Gulacar, O.*[1], Cox, C.[2], Tribble, E.[1], Rothbart, N.[1], & Cohen-Sandler, R.[1]

[1] Department of Chemistry, University of California-Davis, Davis, California, 95616

[2] Department of Chemistry, Duke University, Box 90346, 128 Science Drive Durham, NC 27708-0346

ABSTRACT

The development of problem-solving skills – particularly with stoichiometry concepts – is paramount for succeeding in a general chemistry sequence. Key concepts related to problem solving and stoichiometry were analyzed and reported in this study. The research study analyzed retention of stoichiometry concepts over two consecutive quarters, the correlations between metacognition and success, and the correlations between the COSINE (Coding System for Investigating Subproblems and the Network) codes with course grades and the categories measured by the Metacognitive Awareness Inventory (MAI). Two cohorts, identified as the general and focus group, were evaluated in the study. The general group (N=39) took MAI in fall and winter quarters and completed one multi-step question as a part of their regular exam. Concurrently, the focus group (N=20) participated in a think-aloud session in which they solved six stoichiometry questions. Using a 95% confidence level, statistical differences between the fall and winter problem solving performances were observed with both the general and focus group cohorts. Furthermore, statistically significant correlations (using a 95% level of confidence) were observed between the MAI categories and the COSINE codes.

KEYWORDS

Undergraduate, Problem Solving, Stoichiometry, Metacognition, Knowledge Retention.

INTRODUCTION

Problem solving is a necessary skill for undergraduate and graduate students in STEM courses and research and postgraduates in the workforce. In research and industry alike, employers often seek employees with strong problem-solving capabilities, as new technologies and scientific knowledge depend upon real-world contextual problem solving.[1, 2] Problem solving is an integral component of most STEM courses,[3, 4] especially in general chemistry.[5, 6] Students in general chemistry courses are not only responsible for learning the underlying concepts, they are also required to apply their knowledge to new situations they have not previously encountered.[7-9]

Research into problem solving in the chemistry classroom has been widely reported with studies including: the analysis of cognitive mechanisms for problem solving,[10-12] the development of interventions (teaching strategies) for improving problem solving skills,[13-17] and the role of metacognition in problem solving.[18-23] Of the topics in general chemistry, stoichiometry is widely researched because it is one of the first topics students encounter.[24-26] Stoichiometry skills are embedded in future class concepts, and strategies to solve stoichiometric problems are used throughout general chemistry sequences. For example, gas laws draw upon students' proportional reasoning abilities when comparing pressure and volume, volume and temperature, etc., and incorporate the idea of moles and molecules quantitatively in calculations and qualitatively when describing the kinetic theory of gases. The existing research into stoichiometry problem solving has analyzed and identified relationships between students' proportional reasoning abilities, high school math preparation, and metacognition.[21, 27] This study focuses on subproblem analysis to gain deeper insight into students' problem-solving approaches and strategies while simultaneously exploring how knowledge retention and metacognition relate to this analysis.

Past studies have reported challenges with evaluating students' proficiency with stoichiometry and their problem-solving abilities due to overemphasis of the importance of the end results, rather than the subproblems or steps that lead to students' final answers.[9]

Although some studies focus on students' success with subproblems, they mainly examine whether students got them right or wrong without doing an in-depth investigation of the nature of their mistakes. A carefully planned analysis of the subproblems in stoichiometry questions is necessary to provide greater insight into the successes and challenges students face in answering these questions and thus provide instructors more effective approaches for teaching stoichiometric concepts.

Our research uses the COSINE (Coding System for Investigating Subproblems and the Network) method, reported by Gulacar et al.[28] to carefully analyze subproblems within stoichiometry questions. COSINE (see Table 1) provides an approach to identify the nature and origin of students' challenges in problem solving with specific codes given for each subproblem.

Table 1. COSINE categories and codes with examples

Categories	Codes	Explanation	Example
Successful	Successful (S)	Assigned when student performs subproblem correctly	Student correctly converts grams of a substance into moles of a substance
Neutral	Not Required (NR)	Assigned when student skips a subproblem but uses a different correct method that does not need that subproblem	Student does not find limiting reactant but calculates the number of grams made for each reactant (pretending the rest are in excess), then chooses the smallest gram number
	Did not Know to Do (DD)	Assigned when student skips over a necessary subproblem	Student does not balance chemical equation before doing calculations
	Did Something Else (DSE)	Assigned when student does something different instead of the necessary subproblem	Student adds moles of reactants together instead of finding limiting reactant
Unsuccessful	Unsuccessful-Did Incorrectly (UDI)	Assigned when student understands the needed step but performs it incorrectly	Student makes a math error when converting grams to moles
	Unsuccessful-Received Hint (URH)	Rarely assigned; given when focus group student receives hint from research assistant	Student does not know where to start, so research assistant suggests a step
	Unsuccessful-Guessed (UG)	Assigned when student guesses or it appears that they have guessed	Student writes down an answer with no work preceding it
	Could Not Do (CD)	Assigned when focus group student articulates step needed but cannot do it; also used in the general group for sub steps that students ran out of time to do	Student states that grams must be converted into moles but does not remember to divide by the molar mass

For example, stoichiometry problems can be broken into subproblems that include writing chemical equations, balancing chemical reactions, calculating moles, and identifying the limiting reagent. Eight codes are used with the COSINE method to measure students' success with a subproblem: Successful, Could not Do, Did not Know to Do, Unsuccessful Guessed, Unsuccessful Received Hint, Did Something Else, Unsuccessful Did Incorrectly, and Not Required. The eight codes are further categorized as successful, unsuccessful, or neutral to better evaluate students' success with each subproblem, identify the nuances in students' problem-solving strategies, and provide insight into specific limitations observed regarding unsuccessful problem-solving strategies.

Metacognition, or how actively students think about their own learning and problem-solving strategies, also affects how students learn information and apply what they learn.[29] Studies have indicated that many students, especially those who are lower-performing, tend to overstate their abilities and understanding of content.[30, 31] Additional studies have shown a positive correlation between expertise in a subject and self-awareness of their own thinking processes.[32, 33] Furthermore, a better metacognitive awareness has been correlated with more productive job performance in industry[34] and a better performance in the classroom.[29, 35, 36] This research uses the metacognitive awareness inventory (MAI)[37] which is 52 questions and provides two components of information. The first-component measures students' knowledge about cognition, while the second-component measures students' regulation of cognition abilities. A numerical score of students' metacognitive abilities in each of the categories in both components provide a metric for correlating them with problem solving success.

Knowledge retention refers to students' ability to remember topics and concepts taught in previous courses[38] and is highly desired in educational institutions. For chemistry courses, a cumulative understanding is paramount for success. In general, studies show that knowledge retention decreases when a time gap exists between exposure to the material and its related assessment. A study examining whether learning modules improve students' learning about lipids showed that the scores decreased significantly between a quiz and a test of similar difficulty with only a four-day retention interval between them.[39] Other studies focus on improving students' knowledge retention and researching what can cause these improvements, implying that students are not retaining knowledge as much as desired.[40-42] This study was designed to expand upon current literature by analyzing how continual exposure impacts performance. The qualitative insights into how students solve stoichiometry problems were coded using the COSINE method to provide a metric to assess longitudinal change. The qualitative changes in how students approach stoichiometry subproblems were analyzed using a pre- and a post-test after 14 weeks of exposure to chemical concepts.

METHODOLOGY

Research Questions

The research questions below are posed to analyze the relationships between performance on subproblems of stoichiometry questions, components of metacognition determined by MAI, and overall class performance.

- How do the students' success change over 14-week period with regard to their use of subproblems and the overall problems?
- What correlations exist, if any, between student grades, MAI scores, and COSINE codes?

Participants

Upon the IRB approval, twenty students from the Fall course (CHE 2A), the first course in the General Chemistry series, were invited to participate in a think-aloud session and solve six stoichiometry questions to provide data for in-depth quantitative and qualitative analyses. This cohort was labeled as the focus group. Simultaneously, additional data were collected from general population taking the same course, CHE 2A, in Fall quarter at a research university in Northern California. During the fall course, a problem was included for the analysis in all the students' first midterm. This guaranteed students' participation and encouraged motivation to do well, provided that students wanted to be successful on this test. The main goal for this additional data was to have a high number of students to provide statistical power and reveal hidden relationships that have not been captured in previous studies. In the following quarter, the students registered for taking CHE 2B with the same professor in Winter quarter were asked to complete the same problem on their first midterm as extra credit since the problem was not directly related to the covered materials in that course. Thirty-nine of the chemistry students in these courses were identified eligible to be considered for the general group. These were the only students who took these chemistry courses from the same professor and whose work were available from both quarters. This meant that, though a larger sample was desired, the general group was limited to these 39 students by the necessity of comparison between the two courses. Though some of the focus group students could also be in the general group, it was not investigated who and how many students are in both groups since there was no intention to compare the work of these two groups.

Instruments and Design

The general group was given a single, multi-step stoichiometric question involving common topics in stoichiometry, with subproblem categories including balancing chemical equations (BEQ), mole concept (MC), stoichiometric ratio (SR), and limiting reactant (LR). The time restriction on the general group's problem is difficult to pinpoint, since the problems were given in an individual exam setting and were not the only problems on the test, which was comprised of 24 multiple-choice and 5 free response questions. The students in the focus group were given 60 minutes to complete six problems. These students interviewed separately and were compensated for their time with $15 in school credit during the fall session and $25 more if they participated in the winter session as well. A “think-aloud protocol” was used for the focus group to gain a better understanding about how students think about problems. Students were asked to read the question out loud and talk through the steps they used to solve the given problem. Thus, the exact places where students' thought processes began to go in the wrong direction were easily pinpointed with this protocol, and the reasons for this misdirection were much clearer. Audio and video recordings were taken of the focus group working through their problem sets, but the use of real names on paper or in rosters was avoided. The videos did not capture faces or identifying features. These recordings were then transcribed and used for interpreting students' solutions and analyzing the findings of the study, which was approved by the Institutional Review Board of University of California and assigned the IRBNet ID of 922670-2.

In order to examine students' metacognition, the Metacognitive Awareness Inventory (MAI)[37] was given to the students in the general group in Fall and Winter quarters as an extra credit assignment. Students who completed the MAI also provided information about their grade in CHE 2A, the first course in the series. The MAI includes 52 true/false questions, which were categorized under two components, knowledge of cognition and regulation of cognition. The first component has three subcategories, which are declarative knowledge (8 questions), procedural knowledge (4), and conditional knowledge (5). The second component has the following subcategories: planning (7), information management strategies (10), monitoring (7), debugging strategies (5), and evaluation of learning (6). Students answered these true/false questions based on their perceptions of their study habits and problem-solving skills, and their answers were converted into numbers by counting the number of true answers. For each sub-category, a score was determined and utilized to investigate if any correlation exists between those categories and the codes assigned to the steps, subproblems, in students' solutions of the problems.

Data Analysis

All problems in this study were coded using the Coding System for Investigating Subproblems and Network (COSINE) developed by Gulacar et al.[28] In order to better understand the presented findings in this paper, the reader can also read Gulacar et al.'s[28] paper to get in-depth information regarding the COSINE codes, how the COSINE should be applied in different cases, and how students' solutions and transcripts, if available, could be utilized in that process. This coding system assesses students' success on individual subproblems, while more traditional methods of assessing students' success on a problem rely only on the final answer, which could miss valuable information regarding the exact spots students deviate from the correct solution or get stuck. Each subproblem was assessed using one of the eight codes, and these codes were grouped into three categories: successful, neutral, and unsuccessful, as seen in Table 1. The successful and unsuccessful codes—S, UDI, UG, CD, and URH—give insight into how successful students are on particular subproblems, since the codes in these categories rely on students knowing the next subproblem to be completed. The neutral codes—DD, DSE, and NR—hold importance as indicators of students' success on the problem as a whole, since these entail cases where students do not do the expected subproblems present on the pre-determined rubric.

Because students' work was coded by more than one researcher, Cohen's kappa was used to check for inter-rater reliability. This calculation used the correlation between coding groups (successful, neutral, unsuccessful) instead of the codes themselves (NR, UDI, etc.).[43] Our average kappa value for the Fall data was 0.88, which was deemed acceptable.

The data collected from focus and general groups were employed for different purposes. Due to the high number of students and needed statistical power, the general group data were used for the examination of the relationship between students' problem-solving performance and their metacognitive awareness. On the other hand, the focus group data were utilized to answer the first question that aimed to find out how students' problem-solving success changes over 14-week period.

RESULTS AND DISCUSSION

Investigating underlying challenges in students' problem-solving process by using COSINE

Students' successes with problem solving was investigated using the COSINE codes assigned to each subproblem in students' solutions during both Fall and Winter quarters. To quantify the differences and highlight the influence of knowledge retention and metacognitive awareness on problem-solving performance, the resulting codes were used in two different formulas, Attempt Success Rate (ASR) and Complete Success Rate (CSR). ASR as a percentage represents students' success with a specific subproblem and was determined by dividing successful codes by the sum of successful and unsuccessful codes. Essentially, ASR answers the question, "If a student knows what subproblem is needed, can they execute that task successfully?" CSR was determined in a similar fashion to ASR but also contains the neutral codes DSE (Did Something Else) and DD (Did Not know to Do) in the denominator because these codes indicate that students did not attempt to do the subproblem necessary in the predicted solution. CSR answers the question, "Can a student both determine the subproblem necessary for this step of the overall problem and execute that subproblem successfully?" DD and DSE, in contrast to the other codes, address whether or not a student is addressing the appropriate subproblem, which is why they appear in the CSR formula and not the ASR formula. If students know the exact method of solving a problem, their ASR and CSR scores would be exactly the same, even if a student made math errors or did not know how to execute the subproblems needed; thus the difference between ASR and CSR shows the extent that students do not understand the combination of steps needed to solve a problem. ASR also matters as a measurement of how well a student knows how to solve an individual subproblem, such as converting mass into moles. Even if a student knows the steps needed to solve a problem, it is not helpful if they do not know how to accomplish those individual steps. NR (Not Required) is not included in either the ASR or CSR calculations since it does not affect students' success in any way. It is only applied when a student did not use that specific subproblem in his alternative correct solution. As an example, ASR and CSR for mole concept (MC) were calculated as follows:

$$ASR_{MC} = \frac{S}{S + UDI + UG + URH + CD} x100 \qquad CSR_{MC} = \frac{S}{S + UDI + UG + URH + CD + DD + DSE} x100$$

Both ASR and CSR calculations provided insight into how students' success changed between quarters and additionally allowed easy comparison between participants' success rates. To assess the students' success with problem solving in stoichiometry, the ASR and CSR scores were determined for all the subtopics (WEQ- Writing Equation, BEQ-Balancing Equation, CM-Conversation of Mass, MC-Mole Concept, SR-Stoichiometric Ratio, and LR-Limiting Reagent, PY-Percent Yield, EF-Empirical Formula, MF-Molecular Formula, and MP-Mass Percent) involved in the six problems used during think-aloud protocols in Fall and Winter Quarters (see Figure 1).

The average ASR and CSR values for the fall quarter focus group students in this study were determined to be 78.4% and 74.4%, respectively, whereas the ASR and CSR values for the winter quarter focus group students were 89.9% and 81.2%, respectively. The average ASR scores for both Fall and Winter quarters are higher than CSR scores.

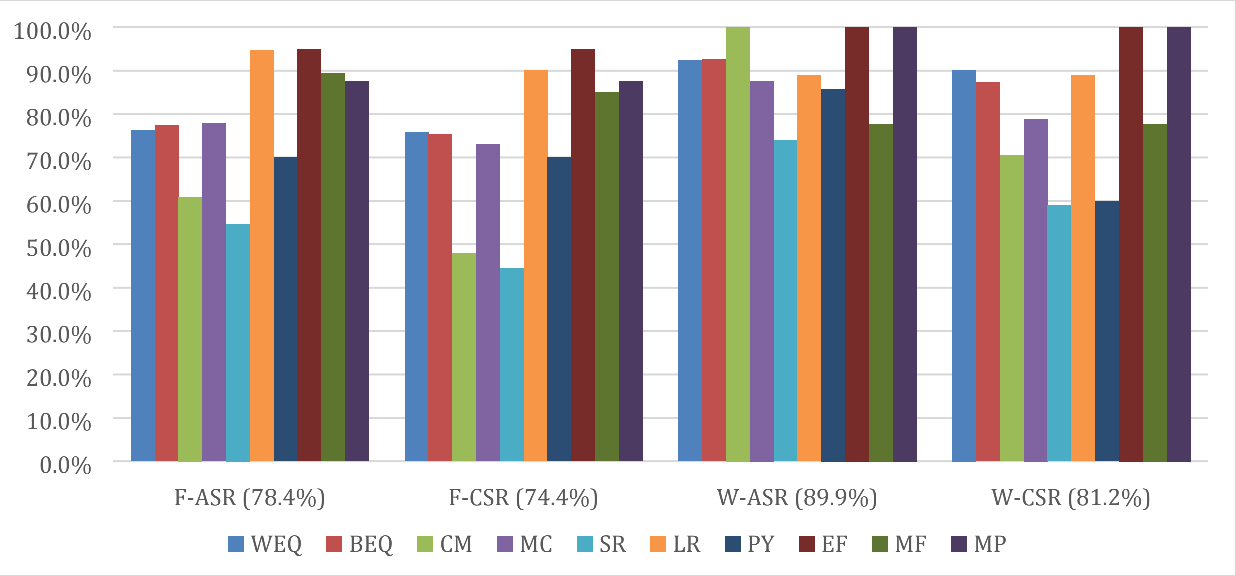


Figure 1. The ASR and CSR scores determined for each subproblem appeared in the problems completed by the focus group students in Fall and Winter quarter.

Although this is expected due to the nature of ASR and CSR calculations, the large differences between the ASR and CSR scores of some of these topics show a large gap between understanding the particular topic and knowing how to use said topics in changing context of problems. Examining the gap between two formulas highlights the source of students' challenges. In these cases, the higher ASR scores still show that the students know those subtopics, and if they know that the solution of the problem requires a subproblem involving one of those subtopics, they execute them successfully. It should be noted that the lower CSR scores are the result of DD (Did Not Know to Do) and DSE (Did Something Else) codes, showing the importance of these two codes in determining students' overall success on a problem. It is speculated that students have high DD and DSE codes when they do not understand what the question is asking or do not recognize the question type. In these cases, they struggle with the question in hand and utilize incorrect strategies that include different irrelevant subproblems, which increases the prevalence of those two codes. When these codes are included in the formula, the scores go down, highlighting how students have greater success at individual subproblems (measured by ASR) than at problems in their entirety (measured by CSR). At least in the context of stoichiometry, the difficulty does not stem from not knowing how to do individual steps involving subtopics in a problem (because the ASR scores were high), it is rather about figuring out what subtopic to use and when. When they are unsure what steps or subtopics are necessary to obtain the correct answer, students start making up steps in solving problems that are not necessarily in the direction of the correct answer. This deviation of their thoughts results in an increased number of DSE or DD codes. The DD and DSE codes are vital because they show the reality of students forgetting necessary steps and deviating from a viable path, respectively.

For the focus group, the majority of DD and DSE codes occurred in the last two questions of the problem set, which can be described as complex problems and were less straightforward than the rest of the problem set. It is a known fact that when students have cognitive overload, they struggle with the task in hand and their performance drops dramatically.[44] It is possible that the number of DD and DSE codes are significantly correlated with cognitive overload. This aspect will be explored more in the future studies. These results indicate that students need more practice with complex questions, rather than doing simple exercises overemphasized in the end-of-chapter questions that assess their understanding of individual subtopics.

The ability to remember concepts and strategies is of high importance in the context of chemistry, where understanding new content relies on a firm understanding of more fundamental concepts.[41] Unfortunately, studies have shown that knowledge retention generally decreases over time for various disciplines.[45, 46] In contrast to this, the data in this study show that students retained the information they had learned in the CHE 2A course in the Fall Quarter and further built upon it during the CHE 2B course in the Winter Quarter. A jump in scores for focus group ASR and CSR scores between quarters as shown in Figure is evidence for greater success when solving general chemistry problems in the CHE 2B course, implying knowledge retention between the courses. The part of the increase in scores could be attributed to the possible improvement of students' general problem-solving skills, but it is difficult to segregate this advancement from the knowledge retention or the gain of new knowledge. ASR score went up by 16.5% and CSR by 12.6%. At 95% confidence interval, ASR for subproblems relating to conservation of mass (t=7.414), mole concept (t=2.961), and stoichiometric ratios (t=2.270) improved at a statistically significant level improving by 39%, 10%, and 19% respectively.

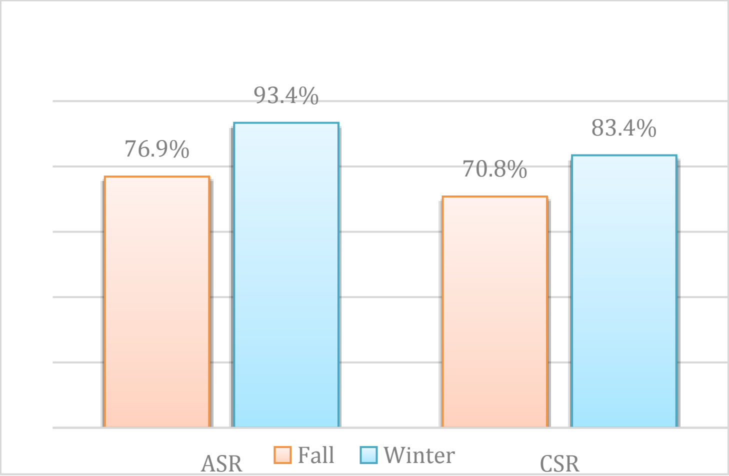


Figure 2. The combined ASR and CSR scores of the focus group for Fall and Winter Quarters

These improvements in problem-solving success may be attributed to the heavy reliance on the fundamental understanding of moles and stoichiometry in CHE 2B to solve problems. This is different from CHE 2A where short, straightforward problems with relatively few steps are tested upon, whereas in CHE 2B many of the entire problems that may appear in CHE 2A are merely part of the process for solving a single, much lengthier and complex CHE 2B problem with many mass and mole conversions amongst new material. These findings were supported with the statistical analysis carried with a bootstrap method as well.

Interestingly, subproblems for balancing equations (BEQ) and limiting reactant (LR) processes failed to show statistically significant improvement between the two quarters. These two types of problems were not emphasized as much in the CHE 2B course and may explain why there is not significant improvement regarding solving these types of problems. There was not a need to practice them outside the basic understanding gained in CHE 2A since the concepts did not evolve much more and stayed reasonably similar. Another reason for the lack of significant difference for these two subproblems could be about their base level. They were measured to be high at the beginning of the study. Therefore, the difference did not appear to be significant although their scores increased from fall to winter session.

Despite increases in both ASR and CSR from Fall to Winter Quarter, the differences between ASR and CSR increased as well. This indicates that the students gained an improved understanding of individual subproblems, but their ability to connect the subproblems to produce the correct solution did not increase at the same rate. Overall, students' abilities to solve both subproblems (ASR) and overall problems (CSR) improved at a statistically significant level ($t = 5.613$, $p<0.05$).

Exploring correlations between metacognitive awareness and problem-solving performance

Students' abilities and performance in science has been linked to their metacognitive abilities in previous research studies .[29, 47] To expand our research to explore and analyze the relationship between students' chemistry problem solving abilities and their metacognitive abilities, students' course grades (A, B, or C) in CHE 2A were correlated with their scores on the MAI sections using the Mann-Whitney test. The nonparametric Mann-Whitney U-tests were used to identify correlations because of the small sample sizes. At a 95% confidence level, a statistically significant correlation between class performance and the MAI declarative knowledge category was observed. Metacognitive abilities with respect to declarative knowledge relate students' abilities to successfully reflect upon their understanding of content and their abilities to apply content in solving problems.[37] Statistically significant correlations between other MAI categories and class performance were not observed.. Table 2. Comparison of MAI inventory subsections and performance for Fall Quarter Analysis(General Gorup) provides a summary of the Mann-Whitney U-values of grades vs. MAI category scores. The lack of a statistically significant correlation with other MAI categories is an interesting finding but can be rationalized based upon the experience level of students in the study. The students analyzed are freshmen who have limited backgrounds in college-level STEM courses, which can account for the wider range of more sophisticated metacognitive abilities such as debugging strategies, evaluation, and planning. Declarative knowledge refers to students' abilities to reflect upon their broad understanding of content. The other metacognitive tasks, in contrast, show students' abilities to not only reflect upon their broad understanding but also their abilities to use, apply, and extend upon their understanding. Additionally, and more importantly, the course assessments which are used to determine final grades are largely multiple choice based, which omits key considerations with how students solve problems. As a result, it becomes more difficult to identify key correlations between specific metacognitive skills and success.

Table 2. Comparison of MAI inventory subsections and performance for Fall Quarter Analysis(General Gorup)

MAI Category	*U (C/B)**	*U (crit)*[a]	*U (A/B)***	*U (crit)*[a]	*U (A/C)****	*U (crit)*[a]
Declarative Knowledge	37.5	21	36.5[b]	45	16.5[b]	26
Procedural Knowledge	41.5	21	69	51	42.5	26
Conditional Knowledge	35.5	21	60.5	51	31	26
Planning	39	21	72	51	37	26
Comprehension and Monitoring	32	21	66	51	29	26
Evaluation	29	21	79	51	31	26
Information Management	34.5	21	83.5	51	40	26
Debugging Strategies	42	21	72	51	44	26

*The U (C/B) represents the calculated U-value for the comparison of students earning C and B grades.

**The U (A/B) represents the calculated U-value for the comparison of students earning A and B grades.

***The U (A/C) represents the calculate U-value for the comparison of students earning A and C grades.

[a] The U (critical) is used to determine whether the U (calculated) is significant. When U (critical) > U (calculated), the difference between the two cohorts is statistically significant according to a 95% level of confidence.

[b] These values are statistically significant using a 95% level of confidence

Because of the challenges with associating final grades with metacognitive abilities, this study additionally sought to analyze correlations between specific COSINE codes and metacognitive abilities. Pearson correlation coefficients were computed based upon students' scores on subcategories and the number of times each code was observed. In contrast to grades, the COSINE codes provide more insight into students' problem-solving approaches and provide a more meaningful analyses of problem-solving strategies. The statistically significant relationships, using a 95% level of confidence interval, are provided in Table 3. Correlation between the COSINE codes and MAI subcategories for Fall Quarter Analysis (General Group). Note the data in Table 3 is based upon the data collected in the fall for both the MAI and stoichiometry analysis. The spread of the winter COSINE codes was more limited, preventing the calculation of meaning correlation coefficients for all codes. To best capture students concurrent metacognitive and problem-solving abilities, only the fall data was used.

Table 3. Correlation between the COSINE codes and MAI subcategories for Fall Quarter Analysis (General Group)

COSINE Code	MAI	Pearson Correlation
DD	Procedural Knowledge	0.393
DSE	Monitoring	-0.393
DSE	Evaluation of Learning	-0.400
UG	Monitoring	-0.432
UG	Debugging Strategies	-0.420

The correlation between procedural knowledge and the DD (Did not Know to Do) codes implies that students with a higher procedure knowledge score on the MAI also were apt to receive more DD code in their problem-solving attempts. A higher procedural knowledge score implies that students' have a stronger ability to compile information to develop more successful strategies. The positive correlation between the procedural knowledge MAI category and the DD code may initially appear perplexing; however, students may have been overconfident in their abilities and skipped steps. This overconfidence could have resulted in the assignment of a DD codes. A limitation to the explanation of this correlation is that the general group data involved the COSINE codes, which did not include the think-aloud protocol. It is difficult to make strong claims without having verbal cues related to students' thinking process before and after those missed subproblems.

The DSE (Did Something Else) has a negative correlation with monitoring and evaluation of learning, which implies that students who have high monitoring or evaluation of learning scores were assigned fewer DSE codes. Monitoring is a measure of one's learning of strategy use, and evaluation is one's ability to measure the effectiveness of a strategy. Therefore, the negative correlation with DSE is not surprising given that higher MAI scores in these categories result in fewer instances in which students evaluate the necessary sub-problems poorly by adding unnecessary steps in overall problem schema. A negative correlation was observed between UG (Unsuccessfully Guessed) and monitoring and debugging strategies. This is again reasonable because a greater metacognitive awareness regarding strategy monitoring and debugging strategies implies that students are more apt to rationalize strategies carefully without randomly guessing the appropriate strategy. These significant correlations between the assigned codes and the MAI categories highlight the additional advantages of the COSINE method as applied to find out students' success with solving stoichiometric problems. Deeper relationships between metacognitive abilities and course grades are not always strongly correlated because relationships are not always readily observed with multiple choice assessments common in large enrollment courses.

LIMITATIONS

Of the approximately1800 students enrolled in general chemistry at the institution where the study was completed, the authors could use the data only from a small subset of 39 students due to the mentioned reasons in the participants section. It is suggested that similar studies are completed within the same semester or quarter with several week gap. Otherwise, it will be difficult to track down students' progress in different sections and gather new data. An additional limitation is the presence of a specific students' demographics at the institution where the research completed. The breakdown of the student population is as follows: 32% Asian/Pacific Islander, 23% Hispanic, 23% White, 17% international students, and 5% others. This demographic makes up a unique student population, and the data collected here cannot be easily generalized and applied to every student group.

CONCLUSION

Stoichiometry is an integral concept in general chemistry that is readily used in several topics including gas laws, colligative properties, and chemical equilibrium. Therefore, to be successful in general chemistry, an understanding of stoichiometry is paramount. Understanding the nature of challenges in students' problem solving in stoichiometry will benefit a large group of students and instructors as it affects STEM students' chemistry self-efficacy at early stage of their education.[48]

Comparisons between the number of COSINE codes students received, their overall grades, and their MAI scores led to insights about the nature of students' problem-solving skills and what factors affect their success. The gap between ASR and CSR scores indicates the importance of the DD (Did not Know to Do) and the DSE (Did Something Else) codes. Despite a general trend in increasing problem-solving performance and greater success with subproblems, students' success with putting all those moving parts together and solving stoichiometry problems as a whole is still relatively low. The investigation of knowledge retention revealed that students' abilities to solve stoichiometric problems improved, most likely due to students' utilization of these stoichiometric concepts in both the first and second quarters, promoting better conceptual understanding and problem-solving abilities. The correlation observed between students' overall grades and students' metacognitive awareness of declarative knowledge provides new ways to predict students' success in the class at an early stage.

The high number of assigned DD and DSE codes point out the importance of scaffolding exercises and gradually introducing a wide range of problems with different difficulty levels to help students gain a better understanding of how to use their acquired knowledge of subproblems in different contexts. Although our findings indicate that students become more successful in solving stoichiometry problems over time, it is not an observed result across all curriculum[49], so it is important that instructors encourage review of previous material in the course and include problems that build on previous material in exams to promote knowledge retention between consecutive chemistry classes. Also, instructors should guide their students to assess their needs metacognitively, take more responsibility of their learning, and understand the vital role of knowledge on cognition and regulation of cognition in the learning process so that students can learn to use their metacognitive resources to decide how to effectively improve their problem-solving abilities.

Due to the small sample population of Fall and Winter cohorts, this study should be reproduced with a larger sample population. In addition to administering surveys, students could be interviewed to gain more insight into students' evaluation of their metacognitive awareness in different areas.

Future research will analyze strategies and success in other chemistry concepts as well as demographic factors such as prior exposure to chemistry, math abilities, and gender. Specifically, future research will analyze problem solving strategies in a broader context beyond stoichiometry with topics like equilibrium, kinetics, and thermodynamics, using the same method of comparing COSINE codes, MAI responses, and course grades. By understanding these relationships, it will be possible to design and implement more effective teaching interventions that will increase students' problem-solving abilities in these areas, which appear in the second half of the general chemistry sequence.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ogulacar@ucdavis.edu

REFERENCES

1.Candy, P. C.; Crebert, R. G., Ivory Tower to Concrete Jungle: The Difficult Transition from the Academy to the Workplace as Learning Environments. *The Journal of Higher Education* **1991,** *62* (5), 570-592.
2.Jonassen, D.; Strobel, J.; Lee, C. B., Everyday Problem Solving in Engineering: Lessons for Engineering Educators. *Journal of Engineering Education* **2006,** *95* (2), 139-151.
3.Kennedy, T. J.; Odell, M. R. L., Engaging Students in STEM Education. *Science Education International* **2014,** *25* (3), 246-258.
4.States, N. L., Next generation science standards for states, by states. Washington, DC : National Academies Press: Washington, DC, 2013.
5.Gabel, D. L.; Bunce, D. M., Research on problem solving: Chemistry. In *Handbook of research on science teaching and learning*, Gabel, D. L., Ed. Macmillan: New York, 1994; p 26.
6.Sawrey, B. A., Concept Learning versus Problem Solving: Revisited. *Journal of Chemical Education* **1990,** *67* (3), 253-54.
7.Bhattacharyya, G.; Bodner, G. M., A cultural approach to problem solving. *Educación Química* **2018,** *16* (2), 222.

8.Bodner, G. M.; Herron, J. D., Problem-Solving in Chemistry. In *Chemical Education: Towards Research-Based Practice*, Gilbert, J. K.; De Jong, O.; Justi, R.; Treagust, D. E.; Van Driel, J. H., Eds. Kluwer: Dordrecht, The Netherlands, 2002; pp 235-266.

9.Nurrenbern, S. C.; Pickering, M., Concept Learning versus Problem Solving: Is There a Difference? *Journal of Chemical Education* **1987,** *64* (6), 508-10.

10.Chan, J. Y. K.; Bauer, C. F., Learning and studying strategies used by general chemistry students with different affective characteristics. *Chemistry Education Research and Practice* **2016,** *17* (4), 675-684.

11.Richards-Babb, M.; Curtis, R.; Ratcliff, B.; Roy, A.; Mikalik, T., General Chemistry Student Attitudes and Success with Use of Online Homework: Traditional-Responsive versus Adaptive-Responsive. *Journal of Chemical Education* **2018,** *95* (5), 691-699.

12.Yuriev, E.; Naidu, S.; Schembri, L. S.; Short, J. L., Scaffolding the development of problem-solving skills in chemistry: guiding novice students out of dead ends and false starts. *Chemistry Education Research and Practice* **2017,** *18* (3), 486-504.

13.Broman, K.; Bernholt, S.; Parchmann, I., Using model-based scaffolds to support students solving context-based chemistry problems. *International Journal of Science Education* **2018,** *40* (10), 1176-1197.

14.Cooper, M. M.; Cox, C. T., Jr.; Nammouz, M.; Case, E.; Stevens, R., An Assessment of the Effect of Collaborative Groups on Students' Problem-Solving Strategies and Abilities. *Journal of Chemical Education* **2008,** *85* (6), 866-872.

15.Frey, R. F.; Fink, A.; Cahill, M. J.; McDaniel, M. A.; Solomon, E. D., Peer-Led Team Learning in General Chemistry I: Interactions with Identity, Academic Preparation, and a Course-Based Intervention. *Journal of Chemical Education* **2018**, <xocs:firstpage xmlns:xocs=""/>.

16.Smith, A. L.; Paddock, J. R.; Vaughan, J. M.; Parkin, D. W., Promoting Nursing Students' Chemistry Success in a Collegiate Active Learning Environment: "if i Have Hope, i Will Try Harder". *Journal of Chemical Education* **2018,** *95* (11), 1929-1938.

17.Talanquer, V., Concept Inventories: Predicting the Wrong Answer May Boost Performance. *Journal of Chemical Education* **2017,** *94* (12), 1805-1810.

18.Casselman, B. L.; Atwood, C. H., Improving General Chemistry Course Performance through Online Homework-Based Metacognitive Training. *Journal of Chemical Education* **2017,** *94* (12), 1811-1821.

19.Cooper, M. M.; Sandi-Urena, S., Design and Validation of an Instrument To Assess Metacognitive Skillfulness in Chemistry Problem Solving. *Journal of Chemical Education* **2009,** *86* (2), 240-245.

20.Herrington, D. G.; Sweeder, R. D., Using Text Messages To Encourage Meaningful Self-Assessment Outside of the Classroom. *Journal of Chemical Education* **2018,** *95* (12), 2148-2154.

21.Mathabathe, K. C.; Potgieter, M., Metacognitive monitoring and learning gain in foundation chemistry. *Chemistry Education Research and Practice* **2014,** *15* (1), 94-104.

22.Parker Siburt, C. J.; MacPhail, R. A.; Bissell, A. N., Developing metacognitive and problem-solving skills through problem manipulation. *Journal of Chemical Education* **2011,** *88* (11), 1489-1495.

23.Parlan, P.; Ibnu, S.; Rahayu, S.; Suharti, S., Effects of the metacognitive learning strategy on the quality of prospective chemistry teacher's scientific explanations. *International Journal of Instruction* **2018,** *11* (4), 673-688.

24.Atkins, P., Chemistry: The great ideas. *Pure and Applied Chemistry* **1999,** *71* (6), 927-929.

25.Cooper, M. M.; Posey, L. A.; Underwood, S. M., Core Ideas and Topics: Building Up or Drilling Down? *Journal of Chemical Education* **2017,** *94* (5), 541-548.

26.Holme, T.; Murphy, K., The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map I: General Chemistry. *Journal of Chemical Education* **2012,** *89* (6), 721-723.

27.Haidar, A. H.; Al Naqabi, A. K., Emiratii high school students' understandings of stoichiometry and the influence of metacognition on their understanding. *Research in Science & Technological Education* **2008,** *26* (2), 215-237.

28.Gulacar, O.; Overton, T. L.; Bowman, C. R.; Fynewever, H., A novel code system for revealing sources of students' difficulties with stoichiometry. *Chemistry Education Research and Practice* **2013,** *14* (4), 507-515.

29.Everson, H. T.; Tobias, S., The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science* **1998,** *26* (1), 65-79.

30.Lindsey, B. A.; Nagel, M. L., Do students know what they know? Exploring the accuracy of students' self-assessments. *Physical Review Special Topics - Physics Education Research* **2015,** *11* (2), 020103.

31.Callender, A. A.; Franco-Watkins, A. M.; Roberts, A. S., Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and Learning* **2016,** *11* (2), 215-235.
32.Zhao, N.; Wardeska, J. G.; McGuire, S. Y.; Cook, E., Metacognition: An Effective Tool to Promote Success in College Science Learning. *Journal of College Science Teaching* **2014,** *43* (4), 48-54.
33.Radovan, M., The Relation between Distance Students' Motivation, Their Use of Learning Strategies, and Academic Success. *Turkish Online Journal of Educational Technology - TOJET* **2011,** *10* (1), 216.
34.Sternberg, R., Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science* **1998,** *26* (1), 127-140.
35.Nelson, T. O.; Dunlosky, J.; Graf, A.; Narens, L., Utilization of Metacognitive Judgments in the Allocation of Study During Multitrial Learning. *Psychological Science* **1994,** *5* (4), 207-213.
36.Miller, T. M.; Geraci, L., Improving metacognitive accuracy: How failing to retrieve practice items reduces overconfidence. *Consciousness and Cognition* **2014,** *29* (C), 131-140.
37.Schraw, G.; Dennison, R. S., Assessing Metacognitive Awareness. *Contemporary Educational Psychology* **1994,** *19* (4), 460-75.
38.Kulak, V. N., Genevieve; Sharma, Rahul, Does the Use of Case-Based Learning Impact the Retention of Key Concepts in Undergraduate Biochemistry? *International Journal of Higher Education* **2017,** *6* (2), 110-120.
39.Horn, S.; Hernick, M., Improving Student Understanding of Lipids Concepts in a Biochemistry Course Using Test-Enhanced Learning. *Chemistry Education Research and Practice* **2015,** *16* (4), 918-928.
40.Lysne, S. J.; Miller, B. G., Research and Teaching: A Comparison of Long-Term Knowledge Retention between Two Teaching Approaches. *Journal of College Science Teaching* **2017,** *46* (6), 100-107.
41.Tucci, V. K.; O'Connor, A. R.; Bradley, L. M., A Three-Year Chemistry Seminar Program Focusing on Career Development Skills. *Journal of Chemical Education* **2014,** *91* (12), 2071-2077.
42.Tamir, P.; Amir, R., High School Science as Viewed by College Students in Israel. *Studies in Educational Evaluation* **1981,** *7* (2), 211-225.
43.Gulacar, O.; Fynewever, H., A Research Methodology for Studying What Makes Some Problems Difficult to Solve. *International Journal of Science Education* **2010,** *32* (16), 2167-2184.
44.Dhlamini, J. J., Enhancing learners' problem solving performance in mathematics: A cognitive load perspective. *European Journal of STEM Education* **2016,** *1* (1), 27-36.
45.Balemans, M. C. M.; Kooloos, J. G. M.; Donders, A. R. T.; Van der Zee, C. E. E. M., Actual Drawing of Histological Images Improves Knowledge Retention. *Anatomical Sciences Education* **2016,** *9* (1), 60-70.
46.Bell, D.; Harless, C.; Higa, J.; Bjork, E.; Bjork, R.; Bazargan, M.; Mangione, C., Knowledge Retention after an Online Tutorial: A Randomized Educational Experiment among Resident Physicians. *Journal of General Internal Medicine* **2008,** *23* (8), 1164-1171.
47.Atmatzidou, S.; Demetriadis, S.; Nika, P., How Does the Degree of Guidance Support Students' Metacognitive and Problem Solving Skills in Educational Robotics? *Journal of Science Education and Technology* **2018,** *27* (1), 70-85.
48.Taasoobshirazi, G.; Glynn, S. M., College students solving chemistry problems: A theoretical model of expertise. *Journal of Research in Science Teaching* **2009,** *46* (10), 1070-1089.
49.Lindsey, R. V.; Shroyer, J. D.; Pashler, H.; Mozer, M. C., Improving Students' Long-Term Knowledge Retention Through Personalized Review. *Psychological Science* **2014,** *25* (3), 639-647.

Investigation of the Correlation Between College Students' Success with Stoichiometry Subproblems and Metacognitive Awareness

Gulacar, O.*[1], Cox, C.[2], Tribble, E.[1], Rothbart, N.[1], & Cohen-Sandler, R.[1]

[1] Department of Chemistry, University of California-Davis, Davis, California, 95616

[2] ~~Department of Chemistry, Stanford University, Stanford, California, 94305~~ Department of Chemistry, Duke University, Box 90346, 128 Science Drive Durham, NC 27708-0346

Abstract

The development of problem-solving skills – particularly with stoichiometry concepts – is paramount for succeeding in a general chemistry sequence. Key concepts related to problem solving and stoichiometry were analyzed and reported in this study. The research study analyzed retention of stoichiometry concepts over two consecutive quarters, the correlations between metacognition and success, and the correlations between the COSINE (Coding System for Investigating Subproblems and the Network) codes with course grades and the categories measured by the Metacognitive Awareness Inventory (MAI). Two cohorts, identified as the general and focus group, were evaluated in the study. ~~Although both groups~~The general group (N=39) took MAI in fall and winter quarters and ~~, the general group (N=39)~~ completed ~~only~~ one multi-step question as a part of their regular exam. ~~Conversely~~Concurrently, t~~T~~he focus group ~~participants~~ (N=20) participated in a think-aloud session in which they solved ~~five~~ six stoichiometry questions. Using a 95% confidence level, statistical differences between the fall and winter problem solving performances were observed with both the general and focus group cohorts. Furthermore, statistically significant correlations (using a 95% level of confidence) were observed between the MAI categories and the COSINE codes.

KEYWORDS
Undergraduate, Problem Solving, Stoichiometry, Metacognition, Knowledge Retention.

Introduction

Problem solving is a necessary skill for undergraduate and graduate students in STEM courses and research and postgraduates in the workforce. In research and industry alike, employers often seek employees with strong problem-solving capabilities, as new technologies and scientific knowledge depend upon real-world contextual problem solving.[1, 2] Problem solving is an integral component of most STEM courses,[3, 4] especially in general chemistry.[5, 6] Students in general chemistry courses are not only responsible for learning the underlying concepts, they are also required to apply their knowledge to new situations they have not previously encountered.[7-9]

Research into problem solving in the chemistry classroom has been widely reported with studies including: the analysis of cognitive mechanisms for problem solving,[10-12] the development of interventions (teaching strategies) for improving problem solving skills,[13-17] and the role of metacognition in problem solving.[18-23] Of the topics in general chemistry, stoichiometry is widely researched because it is one of the first topics students encounter.[24-26] Stoichiometry skills are embedded in future class concepts, and strategies to solve stoichiometric problems are used throughout~~ a~~ general chemistry sequences. ~~For e~~For~~In~~ example, gas laws draw upon students' proportional reasoning abilities when comparing pressure and volume, volume and temperature, etc., and incorporate the idea of moles and molecules~~ in~~ quantitatively in calculations and qualitatively when describing the kinetic theory of gases. The existing research into stoichiometry problem solving has analyzed and identified relationships between students' proportional reasoning abilities, high school math preparation, and metacognition.[21, 27] This study focuses on subproblem analysis to gain deeper insight into students' problem-solving approaches and strategies while simultaneously exploring how knowledge retention and metacognition relate to this analysis.

Past studies have reported challenges with evaluating students' proficiency with stoichiometry and their problem-solving abilities due to overemphasis of the importance of the end results, rather than the subproblems or steps that lead to students' final answers.[9] Although some studies focus on students' success with subproblems, they mainly examine whether students got them right or wrong without doing an in-depth investigation of the nature of their mistakes. ~~Even when subproblems are taken into account, other challenges, such as identifying how and where a student's solution deviated from the correct response.~~ A ~~structured~~ carefully planned analysis of the subproblems in stoichiometry questions is necessary to provide greater insight into the successes and challenges students face in answering these questions and thus provide instructors more effective approaches for teaching stoichiometric concepts.

Our research ~~will use~~uses the COSINE (Coding System for Investigating Subproblems and the Network) method, reported by Gulacar et al.[28] to carefully analyze subproblems within stoichiometry questions. COSINE (~~See~~ see Table 1) provides an approach to identify the nature and origin of students' challenges in problem solving with specific codes given for each subproblem. For example, stoichiometry problems can be ~~problem~~broken into subproblems that include writing chemical equations, balancing chemical reactions, calculating moles, and identifying the limiting reagent. Eight codes are used with the COSINE method to measure students' success with a subproblem: Successful, Could not Do, Did not Know to Do, Unsuccessful Guessed, Unsuccessful Received Hint, Did Something ~~e~~Else, Unsuccessful Did Incorrectly, and Not Required. The eight codes are further categorized as successful, unsuccessful, or neutral to better evaluate students' success with each subproblem, identify the nuances in students' problem-solving strategies, and provide insight into specific limitations observed regarding unsuccessful problem-solving strategies.

Table 1. COSINE categories and codes with examples

Categories	Codes	Explanation	Example
Successful	Successful (S)	Assigned when student performs subproblem correctly	Student correctly converts grams of a substance into moles of a substance
Neutral	Not Required (NR)	Assigned when student skips a subproblem but uses a different correct method that does not need that subproblem	Student does not find limiting reactant but calculates the number of grams made for each reactant (pretending the rest are in excess), then chooses the smallest gram number
	Did not Know to Do (DD)	Assigned when student skips over a necessary subproblem	Student does not balance chemical equation before

			doing calculations
	Did Something Else (DSE)	Assigned when student does something different instead of the necessary subproblem	Student adds moles of reactants together instead of finding limiting reactant
Unsuccessful	Unsuccessful-Did Incorrectly (UDI)	Assigned when student understands the needed step but performs it incorrectly	Student makes a math error when converting grams to moles
	Unsuccessful-Received Hint (URH)	Rarely assigned; given when focus group student receives hint from research assistant	Student does not know where to start, so research assistant suggests a step
	Unsuccessful-Guessed (UG)	Assigned when student guesses or it appears that they have guessed	Student writes down an answer with no work preceding it
	Could Not Do (CD)	Assigned when focus group student articulates step needed but cannot do it; also used in the general group for substeps that students ran out of time to do	Student states that grams must be converted into moles but does not remember to divide by the molar mass

~~Eight codes are used with the COSINE method to measure students' success with a subproblem: Successful, Could not Do, Did not Know to Do, Unsuccessful Guessed, Unsuccessful Received Hint, Did Something else, Unsuccessful Did Incorrectly, and Not Required. These eight codes are further categorized as successful, unsuccessful, or neutral to better evaluate students' success with each subproblem, identify the nuances in students' problem-solving strategies, and provide insight into specific limitations observed regarding unsuccessful problem-solving strategies.~~

Metacognition, or how actively students think about their own learning and ~~problem solving~~problem-solving strategies, also affects how students learn information and apply what they learn.[29] Studies have indicated that many students, especially those who are lower-performing, tend to overstate their abilities and understanding of content.[30, 31]~~.~~ Additional studies have shown a positive correlation between expertise in a subject and self-awareness of their own thinking processes.[32, 33] Furthermore, a better metacognitive awareness has been correlated with more productive job performance in industry[34] and a better performance in the classroom.[29, 35, 36] This research uses the metacognitive awareness inventory (MAI)[37] which is 52 questions ~~and provides~~and provides two components of information. The first-component measures students' knowledge about cognition, while the second-component measures students' regulation of cognition abilities. A numerical score of students' metacognitive abilities in each of the categories in both components provide a metric for correlating them with problem solving success. ~~Therefore, due to this embedded information divided into components, we are using a modified version of the MAI for this study.~~

Knowledge retention refers to students' ability to remember topics and concepts taught in previous courses[38] and is highly desired in educational institutions. For chemistry courses, a cumulative understanding is paramount for success. In general, studies show that knowledge retention decreases when a time gap exists between exposure to the material and its related assessment. A study examining whether learning modules improve students' learning about lipids showed that the scores decreased significantly between a quiz and a test of similar difficulty with only a four-day retention interval between them.[39] Other studies focus on improving students' knowledge retention and researching what can cause these improvements, implying that students are not retaining knowledge as much as desired.[40-42] ~~.~~This study was designed to expand upon current literature by analyzing how continual exposure impacts performance. The qualitative insights into how students solve stoichiometry problems~~,~~ were coded using the COSINE method to provide a metric to assess longitudinal change. The qualitative changes in how students approach stoichiometry subproblems were analyzed using a pre- and a post-test after 14 weeks of exposure to chemical concepts.

METHODOLOGY

Research Questions

The research questions below are posed to analyze the relationships between performance on subproblems of stoichiometry questions, components of metacognition determined by MAI, and overall class performance.

- How do the students' success change over 14-week period with regard to their use of subproblems and the overall problems?

- What correlations exist, if any, between student grades, MAI scores, and COSINE codes?

Participants

Upon the IRB approval, twenty students from the Fall course~~two groups of students, general and focus groups,~~ (CHE 2A), the first course in the General Chemistry series, were invited to participate in ~~the study~~a think-aloud session and solve six stoichiometry questions to provide~~ to collect~~ data for in-depth quantitative and qualitative analyses. This cohort was labeled as the focus group. ~~-~~Simultaneously, additional data were collected from general population taking the same course, CHE 2A, in Fall quarter at a research university in Northern California. During the fall course, a problem was included for the analysis in all the students' first midterm. This guaranteed students' participation and encouraged motivation to do well, provided that students wanted to be successful on this test. The main goal for this additional data was to have a high number of students to ~~that will~~ provide statistical power and reveal hidden relationships that have not been captured in previous studies. In the following quarter, the students registered for taking CHE 2B with the same professor in Winter quarter were asked to complete the same problem on their first midterm as extra credit since the problem was not directly related to the covered materials in that course~~The general group of students (N=39) had completed two consecutive general chemistry courses (CHE 2A and CHE 2B) in fall and winter sessions with the same chemistry professor at a research university in Northern California. During the fall course, a problem was included for the analysis in all the students' first midterm~~. ~~This guaranteed students' participation and encouraged motivation to do well, provided that students wanted to be successful on this test. For the winter course, the same problem was inserted into the students' tests as extra credit since the problem was not directly related to the covered materials in that course.~~ Thirty-nine of the chemistry students in these courses were identified eligible to be considered for ~~the~~the general group. These were the only students who took these chemistry courses from the same professor and whose work were available from both quarters. This meant that, though a larger sample was desired, the general group was limited to these 39 students by the necessity of comparison between the two courses. ~~The second set of students, the focus group (N=20), was selected from the fall general chemistry course for a much more involved problem-solving session. The students in this group were categorized as high-(N=10) and low-achieving (N=10) students based on their first midterm scores in the fall quarter. The students who received a score 70% or higher were considered as high-achievers and the students who scored 40% or lower were grouped as low-achievers. In the winter quarter, five students from both high- and low-achieving groups were invited to re-measure their problem-solving performance by using slightly modified versions of the questions used in the fall quarter.~~ Though some of the focus group~~s~~ students ~~were~~ could also be in the general group, it was not investigated who and how many students are in both groups since there was no intention to ~~no analysis was done~~compare the ~~comparing their~~ work of these two groups.~~for both groups to better preserve the anonymity of the participants.~~

Instruments and Design

The general group was given a single, multi-step stoichiometric question involving common topics in stoichiometry, with subproblem categories including balancing chemical equations (BEQ), mole concept (MC), stoichiometric ratio (SR), and limiting reactant (LR)~~, and others.~~. The time restriction on the general group's problem~~s~~ is difficult to pinpoint, since the problems were given in an individual exam setting and were not the only problems on the test, which was comprised of 24 multiple-choice and 5 free response questions. The students in the focus group were given 60 minutes to complete six problems~~, with an optional seventh problem if time permitted~~. These students interviewed separately and were compensated for their time with $15 in school credit during the fall session and $25 more if they participated in the winter session as well. A “think-aloud protocol” was used for the focus group to gain a better understanding about how students think about problems. Students were asked to read the question out loud and talk through the steps they used to solve the given problem. Thus, the exact places where students' thought processes began to go in the wrong direction were easily pinpointed with this protocol, and the reasons for this misdirection were much clearer. Audio and video recordings were taken of the focus group working through their problem sets, but the use of real names on paper or in rosters was avoided. The videos did not capture faces or identifying features. These recordings were then transcribed and used for interpreting students' solutions and analyzing the findings of the study, which was approved by the Institutional Review Board of University of California and assigned the IRBNet ID of 922670-2.

In order to examine students' metacognition, ~~a modified form of~~ the Metacognitive Awareness Inventory (MAI)[37] was given to the students in ~~both~~ the general ~~and focus~~ group~~s~~ in Fall and Winter quarters as an extra credit assignment. Students who completed the MAI also provided information about their grade in CHE 2A, the first course in the series. ~~Our~~ The MAI includes 52 true/false questions, ~~(versus Shraw and Dennison's 100-point scale between true and false)~~ which were categorized under two components, knowledge of cognition and regulation of cognition. The first component has three subcategories, which are declarative knowledge (8 questions), procedural knowledge (4), and conditional knowledge (5). The second component has the following subcategories: planning (7), information management strategies (10), monitoring (7), debugging strategies (5), and evaluation of learning (6). Students answered these true/false questions based on their perceptions of their study habits and problem-solving skills, and their answers were converted into numbers by counting the number of true answers. For each sub-category, a score was determined and utilized to investigate if any correlation exists between those categories and the codes assigned to the steps, subproblems, in students' solutions of the problems.

Commented [CC1]: Did we do anything with the winter MAI? I dont recall analyzing any data associated with the winter MAI. I think the fall data was the only data that I used for making the correlations. We would expect students stoichiometry experience to evolve and their metacognitive experiences evolve ... but perhaps not linearly. I need to address this point for the reviewer and that is the strongest argument that I have to effectively address their concerns.

Commented [OG2R2]: I do not think we did anything with Winter MAI data. The file is in the folder.

Data Analysis

All problems in this study were coded using the Coding System for Investigating Subproblems and Network (COSINE) developed by Gulacar et al.[28] In order to better understand the presented findings in this paper, the reader can also read Gulacar et al.'s[28] paper to get~~-~~ in-depth information~~-~~ regarding the COSINE codes, ~~and~~ how the COSINE should be applied in different cases, and ~~and~~ how students' solutions and transcripts, if available, could be utilized in that process. This coding system ~~grades~~assesses students' success on individual subproblems, while more traditional methods of ~~determining a~~ assessing students' success~~es~~

on a problem rely only on the final answer, which could miss valuable information ~~about~~ regarding the exact spots ~~a student~~ students deviate~~s~~ from the correct solution or get~~s~~ stuck. Each subproblem was ~~graded~~assessed using one of the eight codes, and these codes were grouped into three categories: successful, neutral, and unsuccessful, as seen in Table 1. The successful and unsuccessful codes—S, UDI, UG, CD, and URH—give insight into how successful students are on particular subproblems, since the codes in these categories rely on students knowing the next subproblem to be completed. The neutral codes—DD, DSE, and NR—hold importance as indicators of students' success on the problem as a whole, since these entail cases where students do not do the ~~appropriate~~ expected subproblems present on the pre-determined rubric.

Because ~~tThe problems~~students' work ~~were~~was coded by more than one researcher, ~~so~~ Cohen's kappa was used to check for inter-rater reliability. This calculation used the correlation between coding groups (successful, neutral, unsuccessful) instead of the codes themselves (NR, UDI, etc.).[43] Our average kappa value for the ~~fall general group~~Fall data was 0.88, which was deemed acceptable.

The data collected from focus and general groups were employed for different purposes. Due to the high number of students and needed statistical power, the general group data were used for the examination of the relationship between students' problem-solving performance and their metacognitive awareness. On the other hand, the focus group data were utilized to answer the first question that aimed to find out how students' problem-solving success changes over 14-week period.

Commented [OG3]: Charlie, please check if this makes sense.

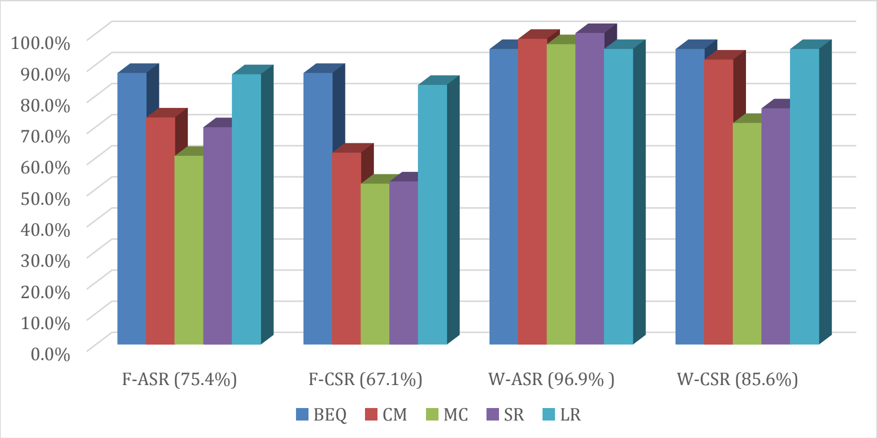
RESULTS AND DISCUSSION

Investigating underlying challenges in students' problem-solving process by using COSINE

Students' successes with problem solving was investigated ~~by utilizing~~using the COSINE codes assigned to each sub~~-~~problem in students' solutions ~~gathered in~~ during both Fall and Winter quarters. To~~In order to~~ quantify the differences and highlight the influence of knowledge retention and metacognitive awareness on problem-solving performance, the resulting codes were used in two different formulas, Attempt Success Rate (ASR) and Complete Success Rate (CSR). ASR as a percentage represents students' success with a specific sub~~-~~problem and was determined by dividing successful codes by the sum of successful and unsuccessful codes. Essentially, ASR answers the question, "If a student knows what subproblem is needed, can they execute that task successfully?" CSR was ~~also~~ determined in a similar fashion to ASR but also contains the neutral codes DSE (Did Something Else) and DD (Did Not know to Do) in the denominator because these codes indicate that students did not attempt to do the subproblem necessary in the predicted solution. CSR answers the question, "Can a student both determine the subproblem necessary for this step of the overall problem and execute that subproblem successfully?" DD and DSE, in contrast to the other codes, address whether or not a student is addressing the appropriate subproblem, which is why they appear in the CSR formula and not the ASR formula. If students know the exact method of solving a problem, their ASR and CSR scores would be exactly the same, even if a student made math errors or did not know how to execute the subproblems needed; thus the difference between ASR and CSR shows the extent that students do not understand the combination of steps needed to solve a problem. ASR also matters as a measurement of how well a student knows how to solve an individual subproblem, such as converting mass into moles. Even if a student knows the steps needed to solve a problem, it is not helpful if they do not know how to accomplish those individual steps. ~~These neutral codes do not reveal any information on students' ability of doing the subproblem evaluated. Therefore, they were not included in the Attempt Success Rate formula.~~ NR (Not Required) is not included in either the ASR or CSR calculations ~~formula~~ since it does not affect students' success in any way. It is only ~~a code~~ applied when a student did not use that specific subproblem in his alternative correct solution. As an~~For~~ example, ASR and CSR for mole concept (MC) were calculated as follows:

$$ASR_{MC} = \frac{S}{S + UDI + UG + URH + CD} x100 \qquad CSR_{MC} = \frac{S}{S + UDI + UG + URH + CD + DD + DSE} x100$$

Both ASR and CSR calculations provided insight into how students' success changed between quarters and additionally allowed easy comparison between participants' success rates. To assess the ~~general group~~ students' success with problem solving in stoichiometry, t~~T~~he ASR and CSR scores were determined for all the subtopics (WEQ- Writing Equation, BEQ-Balancing Equation, CM- Conversation of Mass, MC-Mole Concept, SR-Stoichiometric Ratio, and LR-Limiting Reagent, PY-Percent Yield, EF-Empirical Formula, MF-Molecular Formula, and MP-Mass Percent) involved in the ~~single problem inserted into the midterm exams s~~six problems used during think-aloud protocols in Fall and Winter Quarters (see Figure 1). ~~to assess the general group students' success with problem solving in stoichiometry￼Figure ￼).~~



~~Figure 1. The ASR and CSR scores determined for each subproblem appeared in the problems completed by the focus group students in Fall and Winter quarter.~~

~~The average ASR and CSR values for the fall quarter general group students in this study were determined to be 75.4% and 67.1% respectively, whereas the ASR and CSR values for the winter~~

~~quarter general group students were 96.9% and 85.6% respectively.~~ The average ASR and CSR values for the fall quarter focus group students in this study were determined to be 78.4% and 74.4%, respectively, whereas the ASR and CSR values for the winter quarter focus group students were 89.9% and 81.2%, respectively. ~~The same calculations were repeated to determine ASR and CSR scores for the subtopics involved in the problem set utilized during think-aloud protocols to investigate focus group students' problem-solving performance in both Fall and Winter quarters. In addition to the topics involved in general group's data, as depicted in Figure 2, the focus group data included scores for WEQ-Writing Equation, PY-Percent Yield, EF-Empirical Formula, MF-Molecular Formula, and MP-Mass Percent. The results showed a similar trend for the focus group students.~~ The average ASR scores for both Fall and Winter quarters are higher than CSR scores. Although this is expected due to the nature of ASR and CSR calculations, the large differences between the ASR and CSR scores of some of these topics show a large gap between understanding the particular topic and knowing how to use said topics in ~~the~~changing context of ~~a larger~~ problems.~~.~~ Examining the gap between two formulas highlights the source of students' challenges. In these cases, t~~T~~he higher ASR scores still show that the students ~~do~~ know those subtopics, and if they know that the solution of the problem requires a subproblem involving one of those subtopics, they execute them successfully. It should be noted that the lower CSR scores are the result~~s~~ of DD (Did Not Know to Do) and DSE (Did Something Else) codes, showing the importance of these two codes in determining students' overall success on a problem. It is speculated that students have high DD and DSE codes when they do not understand what the question is asking or do not recognize the question type. In these cases, they struggle with the question in hand and utilize incorrect strategies that include different irrelevant subproblems, which increases the prevalence of those two codes.

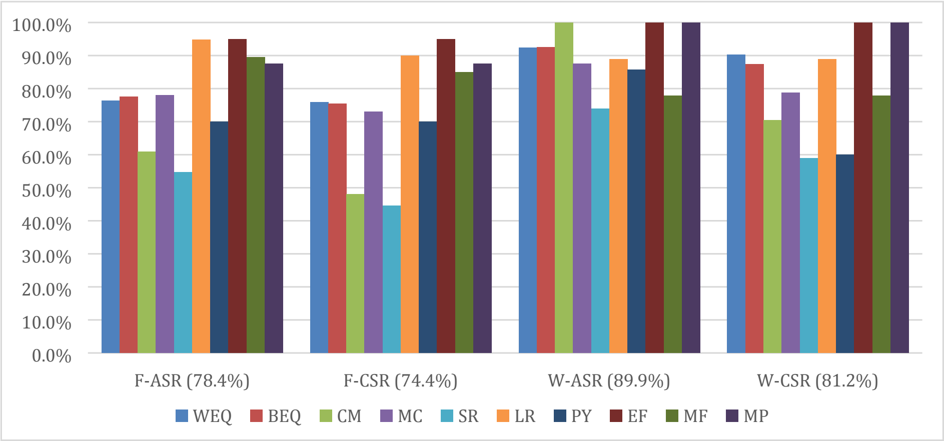


Figure ~~2~~1. The ASR and CSR scores determined for each subproblem appeared in the problems completed by the focus group students in Fall and Winter quarter.

When these codes are included in the formula, the scores go down, highlighting how students have greater success at individual subproblems (measured by ASR) than at problems in their entirety (measured by CSR). At least in the context of stoichiometry, the difficulty does not stem from not knowing how to do individual steps involving subtopics in a problem (because the ASR scores were high), it is rather about figuring out what subtopic to use and when. When they are unsure what steps or subtopics are necessary to obtain the correct answer, students start making up steps in solving problems that are not necessarily in the direction of the correct answer. This deviation of their thoughts results in an increased number of DSE or DD codes. The DD and DSE codes are vital because they show the reality of students forgetting necessary steps and deviating from a viable path, respectively.

~~The greater insight into the focus group's thought processes through the think-aloud protocol changed the numbers of DD and DSE codes but did not eliminate the gap between their ASR and CSR. In Figure 2~~Figure 1~~, the differences between Fall-ASR and Fall-CSR and Winter-ASR and Winter-CSR were recorded as 4.00% and 8.70%, respectively.~~ For the focus group, the majority of DD and DSE codes occurred in the last two questions of the problem set, which can be described as complex problems and were less straightforward than the rest of the problem set. It is a known fact that when students have cognitive overload, they struggle with the task in hand and their performance drops dramatically.[44] It is possible that the number of DD and DSE codes are significantly correlated with cognitive overload. This aspect will be explored more in the future studies. These results indicate that students need more practice with complex questions, rather than doing simple exercises overemphasized in the end-of-chapter questions that assess their understanding of individual subtopics.

The ability to remember concepts and strategies is of high importance in the context of chemistry, where understanding new content relies on a firm understanding of more fundamental concepts.[41] Unfortunately, studies have shown that knowledge retention generally decreases over time for various disciplines.[45, 46]

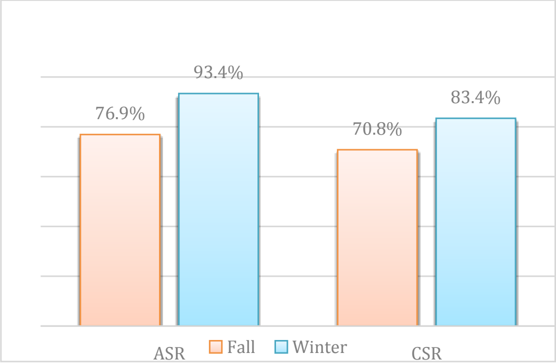


Figure ~~3~~2. The combined ASR and CSR scores of ~~general and~~ the focus group~~s~~ for Fall and Winter Quarters

In contrast to this, the data in this study show~~s~~ that students retained the information they had learned in the CHE 2A course in the Fall Quarter and further built upon it during the CHE 2B course in the Winter Quarter. A jump in scores for ~~both the combined general and~~ focus group ASR and CSR scores between quarters as shown in Figure ~~3~~ is evidence for greater success when solving general chemistry problems in the CHE 2B course, implying knowledge retention between the courses. The part of the increase in scores could be attributed to the possible improvement of students' general problem-solving skills, but it is difficult to segregate this advancement from the knowledge retention or the gain of new knowledge. ASR score went up by 16.5% and CSR by 12.6%. At 95% confidence interval, ASR for subproblems relating to conservation of mass (t=7.414), mole concept (t=2.961), and stoichiometric ratios (t=2.270) improved at a statistically significant level improving by 39%, 10%, and 19% respectively. These improvements in problem-solving success may be attributed to the heavy reliance on the fundamental understanding of moles and stoichiometry in CHE 2B to solve problems. This is different from CHE 2A where short, straightforward problems with relatively few steps are tested upon, whereas in CHE 2B many of the entire problems that may appear in CHE 2A are merely part of the process for solving a single, much lengthier and complex CHE 2B problem with many mass and mole conversions amongst new material. These findings were supported with the statistical analysis carried with a bootstrap method as well.

Interestingly, subproblems for balancing equations (BEQ) and limiting reactant (LR) processes failed to show statistically significant improvement between the two quarters. These two types of problems were not emphasized as much in the CHE 2B course and may explain why there is not significant improvement regarding solving these types of problems. There was not a need to practice them outside the basic understanding gained in CHE 2A since the concepts did not evolve much more and stayed reasonably similar. Another reason for the lack of significant difference for these two subproblems could be about their base level. They were measured to be high at the beginning of the study. Therefore, the difference did not appear to be significant although their scores increased from fall to winter session.

Despite increases in both ASR and CSR from Fall to Winter Quarter, the differences between ASR and CSR increased as well. This indicates that the students gained an improved understanding of individual subproblems, but their ability to connect the subproblems to produce the correct solution did not increase at the same rate. Overall, students' abilities to solve both subproblems (ASR) and overall problems (CSR) improved at a statistically significant level ($t = 5.613$, $p<0.05$).

Commented [CC4]: Perhaps we should elaborate on what this means. I know it means the overall ability but the way it is worded we might need to elaborate to make this more clear.

Commented [OG5R5]: If you do not strongly believe it is necessary, we should not worry about it as there is no specific comment stressing on this point.

Exploring correlations between metacognitive awareness and problem-solving performance

Students' abilities and performance in science has been linked to their metacognitive abilities in previous research studies ~~Metacognition is one of the important factors that~~

~~influences one's performance and ability in their studies and academic performance~~.[29, 47] ~~In order to better~~To expand our research to ~~explore and understand~~explore and analyze the relationship between studen~~ts' chemistry problem solving abilities and their metacognitive abilities~~ts' chemistry problem solving abilities and their metacognitive abilities ~~ts' problem-solving performances~~, students' ~~overall~~ course grades (A, B, or C) ~~they earned~~ in CHE 2A were correlated~~,~~ with their ~~and the~~ scores on the MAI sections using the~~, a~~ Mann-Whitney test. ~~was used to compare groups based upon the final grade (A, B, or C) earned in the course.~~ The nonparametric Mann-Whitney U-tests~~Nonparametric tests~~ were used to ~~compare the samples~~identify correlations because of the small sample sizes. ~~Two key observations emerged when analyzing the correlations between the MAI descriptors and student performance. A statistically significant difference~~ ~~Using~~At a 95% confidence~~l~~ level ~~of confidence,~~ ~~(p < 0.05)~~ a statistically significant correlation between class performance ~~(using letter grades)~~ and the MAI declarative knowledge category ~~in MAI~~ was observed. ~~This indicates that students' metacognition regarding their declarative knowledge their overall performance in the course are strongly correlated.~~ Metacognitive abilities with respect to d~~D~~eclarative knowledge ~~refers to~~relate ~~knowledge~~students' abilities to successfully reflect upon their understanding of content and their abilities to apply content in solving problems. ~~about problem-solving strategies, as well as a student's awareness about how well they learn them~~.[37] ~~There were no other observed statistically significant differences~~Statistically significant correlations between ~~between~~ other~~the~~ MAI categories ~~and~~ and class performance were not observed. ~~performance~~. Table 2. Comparison of MAI inventory subsections and performance for Fall Quarter Analysis (General Gorup)~~(N = 39)~~ provides a summary of the Mann-Whitney U-values of grades vs. MAI category scores.

Table 2. Comparison of MAI inventory subsections and performance for Fall Quarter Analysis (General Gorup)~~(N = 39)~~

MAI Category	*U (C/B)**	*U (crit)*[a]	*U (A/B)***	*U (crit)*[a]	*U (A/C)****	*U (crit)*[a]
Declarative Knowledge	37.5	21	36.5[b]	45	16.5[b]	26
Procedural Knowledge	41.5	21	69	51	42.5	26
Conditional Knowledge	35.5	21	60.5	51	31	26

Planning	39	21	72	51	37	26
Comprehension and Monitoring	32	21	66	51	29	26
Evaluation	29	21	79	51	31	26
Information Management	34.5	21	83.5	51	40	26
Debugging Strategies	42	21	72	51	44	26

*The U (C/B) represents the calculated U-value for the comparison of students earning C and B grades.

**The U (A/B) represents the calculated U-value for the comparison of students earning A and B grades.

***The U (A/C) represents the calculate U-value for the comparison of students earning A and C grades.

[a] The U (critical) is used to determine whether the U (calculated) is significant. When U (critical) > U (calculated), the difference between the two cohorts is statistically significant according to a 95% level of confidence.

[b] These values are statistically ~~significant ones.~~~~sigificant~~significant using a 95% level of confidence

~~While problem solving and metacognitive skills are reflected in course grades, there are several types of qualitative and quantitative problems and other factors that collectively contribute to grades. Therefore, it is not surprisingly that statistical differences were not observed with several of the MAI categories.~~ The lack of a statistically significant correlation with other MAI categories is an interesting ~~finding, but~~finding but can be rationalized based upon the experience level of students in the study. The students analyzed are freshme~~a~~n who have limited backgrounds in college-level STEM courses, which can account for the wider range of more sophisticated metacognitive abilities such as debugging strategies, evaluation, and planning. Declarative knowledge refers to~~is~~ students' abilities to~~ies to~~that reflect upon their broad understanding of content. ~~compared to~~ T~~t~~he other metacognitive tasks, in contrast,~~ which~~ ~~reflect~~show students' abilities to not only reflect upon their broad understanding but also their abilities to use, apply, and extend upon their understanding. Additionally, and more importantly, the course assessments which are used to determine final grades are largely multiple choice based, which omits key considerations with how students solve problems. As a result, it becomes more difficult to identify~~, which makes identifying~~ key correlations between specific metacognitive skills and success~~ more challenging~~.

Commented [EKT6]: This part is readable, but I'm a bit confused at the points being made. Why does the students being freshmen correlate with them having a wider range of metacognitive abilities?

Commented [OG7R7]: Charlie, can we put a reference here?

Commented [CC8]: I think I am going to elaborate on this part and explain why we are not surprised that other relationships were not observed with the grades and the MAI scores. I think this would be a key part to contribute to the quality of the paper. I think the other parts of the MAI are more content focused than declarative knowledge. I am going to think about how to best approach this explanation in order to drawn meaningful conclusions.

~~In addition to checking the correlations between the students' MAI category scores and their overall course grades, the associations between the codes assigned to the subproblems in students' solutions and the categories of MAI were examined.~~ Because of the challenges with associating final grades with metacognitive abilities, this study additionally sought to analyze correlations between specific COSINE codes and metacognitive abilities. Pearson correlation coefficients were computed based upon students' scores on subcategories and the number of times each code was observed. In contrast to grades, t~~T~~he COSINE codes provide more insight into students' ~~problem solving~~problem-solving approaches and provide a more meaningful analyses of problem-~~-~~solving strategies. The statistically significant relationships, using a 95% level of confidence interval, are provided in Table 3. Correlation between the COSINE codes and MAI subcategories for Fall Quarter Analysis (General Group) ~~(N = 39).~~. Note the data in Table 3 is based upon the data collected in the fall for both the MAI and stoichiometry analysis. The spread of the winter COSINE codes was more limited, preventing the calculation of meaning correlation coefficients for all codes. To best capture students concurrent metacognitive and problem solving abilities, only the fall data was used.

Table 3. Correlation between the COSINE codes and MAI subcategories for Fall Quarter Analysis (General Group) ~~(N = 39).~~

COSINE Code	MAI	Pearson Correlation
DD	Procedural Knowledge	0.393
DSE	Monitoring	-0.393
DSE	Evaluation of Learning	-0.400
UG	Monitoring	-0.432
UG	Debugging Strategies	-0.420

~~The first correlation noted indicate that a higher procedural knowledge score indicates a higher likelihood~~The correlation between procedural knowledge and the DD (Did not Know to Do) codes implies that students with a higher procedure knowledge score on the MAI also were apt to receive ~~the~~more DD code ~~in the problem solving schema~~in their problem-solving attempts. A higher procedural knowledge score implies that students' have a stronger ability to compile information to develop more successful strategies. The positive correlation between the procedural knowledge MAI category and the DD code may initially appear perplexing; however, ~~students with stronger content and problem-solving abilities can more readily chunk information~~. ~~Additionally,~~ students may have been ~~overconfidant~~overconfident in~~of~~ their abilities and ~~also~~ skipped steps. This overconfidence could have~~e chunking of information or overconfidence may have~~ resulted in the assignment of a DD codes~~ because students skipped steps~~. A limitation ~~of~~to the explanation of this correlation is that ~~The COSINE method was applied to exam problems to make this correlation, which is limitation because it~~the general group data ~~-~~involved the COSINE codes~~ from the exam problems~~, which did not include the think-aloud protocol. It is difficult to make strong claims without having verbal cues related to students' thinking process before and after those missed subproblems~~This gave verbal cues to ensure more accurate coding, especially in cases where students verbalized a substep without writing it down~~. ~~verbal cues present during think aloud protocols were not present. to receive the DD (Did not Know to Do) code. This could be a result of overconfidence of students when it comes to scoring their ability to generate a strategy to solve exercises and problems. Therefore, students may not necessarily show every step, accounting for the DD code.~~

Commented [OG9]: It sounds chunking is a bad thing. Next sentence is what we should say. I am not sure about this claim.

The DSE (Did Something Else) has a negative correlation with monitoring and evaluation of learning~~. It means a student with a~~, which implies that students who have high monitoring or evaluation of learning scores were assigned fewer~~would receive a lower number of~~ DSE codes~~ or vice versa.~~. Monitoring is a measure of one's learning of strategy use, and evaluation is one's ability to measure the effectiveness of a strategy. Therefore, the negative correlation with DSE is not surprising given that higher MAI scores in these categories ~~would result~~result in ~~a lower number of~~fewer instances in which students evaluate the necessary sub~~-steps~~-problems ~~poorly to get the correct answer and produce irrelevant~~ poorly by adding ~~and~~ unnecessary steps in overall problem schema. A negative correlation was observed between UG (Unsuccessfully Guessed) and monitoring and debugging strategies. ~~This is~~This is again reasonable because a greater metacognitive awareness regarding strategy monitoring and debugging strategies ~~indicate that~~ implies that students are more apt to ~~identify a reasonable strategy without~~

~~needing to guess at the answer~~rationalize strategies carefully without randomly guessing the appropriate strategy. These significant correlations between the assigned codes and the MAI categories highlight the additional advantages of the ~~COSINE for the analysis of students' challenges with problem solving and shedding light on the common type of the mistakes they are making in this process.~~ COSINE method as applied to find out students' success with ~~problem~~ solving ~~and success in~~ stoichiometric~~y~~ problems. Deeper relationships between metacognitive abilities and course grades are not always strongly correlated because relationships are not always readily observed with multiple choice assessments common in large enrollment courses.

Commented [OG10]: We already said it earlier. I guess we can delete it.

LIMITATIONS

Of the approximately1800 students enrolled in general chemistry at the institution where the study was completed, the authors could use the data only from ~~-~~a small subset of 39 students~~-~~ due to the mentioned reasons in the participants section~~was analyzed in the study~~. It is suggested that similar studies are completed within the same semester or quarter with several week gap. Otherwise, it will be difficult to track down students' progress in different sections and gather new data. ~~The application of the COSINE method is time consuming, which limits the reliable application of the method reliably. Future studies will focus upon expanding the subset of student data, and the development of research approaches for analyzing larger sets of data efficiently. The future studies include the application of the COSINE method to other topics in general chemistry to identify whether similar MAI correlations are observed and to further identify challenges students face with problems across the chemistry curriculum. An abbreviated MAI[32] was used for the analysis, and future studies could reveal additional relationships that might emerge from the use of the full version of the inventory—particularly with respect to the items in which statistical relationships were not observed in this study.~~ An additional limitation is the presence of a specific students' demographics at the institution where the research completed. The breakdown of the student population is as follows: 32% Asian/Pacific Islander, 23% Hispanic, 23% White, 17% international students, and 5% others.

Commented [EKT11]: Wait… we expand this idea in the conclusion. Do we need both?

This demographic makes up a unique student population, and the data collected here cannot be easily generalized and applied to every student group.

CONCLUSION

Stoichiometry is an integral concept in general chemistry that is readily used in several topics including gas laws, colligative properties, and chemical equilibrium. Therefore, to be successful in general chemistry, an understanding of stoichiometry is paramount. Understanding the nature of challenges in students' problem solving in stoichiometry will benefit a large group of students and instructors as it affects STEM students' chemistry self-efficacy at early stage of their education.[48]

Comparisons between the number of COSINE codes students received, their overall grades, and their MAI scores led to insights about the nature of students' problem-solving skills and what factors affect their success. The gap between ASR and CSR scores ~~for both the general and the focus group~~ indicates the importance of the DD (Did not Know to Do) and the DSE (Did Something Else) codes. Despite a general trend in increasing problem-solving performance and greater success with sub~~-~~problems, students' success with putting all those moving parts together and solving stoichiometry problems as a whole is still relatively low. The investigation of knowledge retention revealed that students' abilities to solve stoichiometric problems improved~~ in both the general and focus groups~~, most likely due to students' utilization of these stoichiometric concepts in both the first and second quarters, promoting better conceptual understanding and problem-solving abilities. The correlation observed between students' overall grades and students' metacognitive awareness of declarative knowledge provides new ways to predict students' success in the class at an early stage.

The high number of assigned DD and DSE codes point out the importance of scaffolding exercises and gradually introducing a wide range of problems with different difficulty levels to help students gain a better understanding of how to use their acquired knowledge of subproblems in different contexts. Although our findings indicate that students become more successful in solving stoichiometry problems over time, it is not an observed result across all curriculum[49], so it is important that instructors encourage review of previous material in the course and include problems that build on previous material in exams to promote knowledge retention between consecutive chemistry classes. Also, instructors should guide their students to assess their needs metacognitively, take more responsibility of their learning, and understand the vital role of knowledge on cognition and regulation of cognition in the learning process so that students can learn to use their metacognitive resources to decide how to effectively improve their problem-solving abilities.

Due to the small sample population of Fall and Winter cohorts, this study should be reproduced with a larger sample population. In addition to administering surveys, students could be interviewed to gain more insight into students' evaluation of their metacognitive awareness in different areas.

Future research will analyze strategies and success in other chemistry concepts as well as demographic factors such as prior exposure to chemistry, math abilities, and gender. Specifically, future research will analyze problem solving strategies ~~with~~ in a broader context beyond stoichiometry with topics like equilibrium, kinetics, and thermodynamics, using the same method of comparing COSINE codes, MAI responses, and course grades. By understanding these relationships, it will be possible to design and implement more effective teaching interventions that will increase students' problem-solving abilities in these areas, which appear in the second half of the general chemistry sequence.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ogulacar@ucdavis.edu

REFERENCES

1.Candy, P. C.; Crebert, R. G., Ivory Tower to Concrete Jungle: The Difficult Transition from the Academy to the Workplace as Learning Environments. *The Journal of Higher Education* **1991,** *62* (5), 570-592.
2.Jonassen, D.; Strobel, J.; Lee, C. B., Everyday Problem Solving in Engineering: Lessons for Engineering Educators. *Journal of Engineering Education* **2006,** *95* (2), 139-151.
3.Kennedy, T. J.; Odell, M. R. L., Engaging Students in STEM Education. *Science Education International* **2014,** *25* (3), 246-258.
4.States, N. L., Next generation science standards for states, by states. Washington, DC : National Academies Press: Washington, DC, 2013.

5.Gabel, D. L.; Bunce, D. M., Research on problem solving: Chemistry. In *Handbook of research on science teaching and learning*, Gabel, D. L., Ed. Macmillan: New York, 1994; p 26.
6.Sawrey, B. A., Concept Learning versus Problem Solving: Revisited. *Journal of Chemical Education* **1990,** *67* (3), 253-54.
7.Bhattacharyya, G.; Bodner, G. M., A cultural approach to problem solving. *Educación Química* **2018,** *16* (2), 222.
8.Bodner, G. M.; Herron, J. D., Problem-Solving in Chemistry. In *Chemical Education: Towards Research-Based Practice*, Gilbert, J. K.; De Jong, O.; Justi, R.; Treagust, D. E.; Van Driel, J. H., Eds. Kluwer: Dordrecht, The Netherlands, 2002; pp 235-266.
9.Nurrenbern, S. C.; Pickering, M., Concept Learning versus Problem Solving: Is There a Difference? *Journal of Chemical Education* **1987,** *64* (6), 508-10.
10.Chan, J. Y. K.; Bauer, C. F., Learning and studying strategies used by general chemistry students with different affective characteristics. *Chemistry Education Research and Practice* **2016,** *17* (4), 675-684.
11.Richards-Babb, M.; Curtis, R.; Ratcliff, B.; Roy, A.; Mikalik, T., General Chemistry Student Attitudes and Success with Use of Online Homework: Traditional-Responsive versus Adaptive-Responsive. *Journal of Chemical Education* **2018,** *95* (5), 691-699.
12.Yuriev, E.; Naidu, S.; Schembri, L. S.; Short, J. L., Scaffolding the development of problem-solving skills in chemistry: guiding novice students out of dead ends and false starts. *Chemistry Education Research and Practice* **2017,** *18* (3), 486-504.
13.Broman, K.; Bernholt, S.; Parchmann, I., Using model-based scaffolds to support students solving context-based chemistry problems. *International Journal of Science Education* **2018,** *40* (10), 1176-1197.
14.Cooper, M. M.; Cox, C. T., Jr.; Nammouz, M.; Case, E.; Stevens, R., An Assessment of the Effect of Collaborative Groups on Students' Problem-Solving Strategies and Abilities. *Journal of Chemical Education* **2008,** *85* (6), 866-872.
15.Frey, R. F.; Fink, A.; Cahill, M. J.; McDaniel, M. A.; Solomon, E. D., Peer-Led Team Learning in General Chemistry I: Interactions with Identity, Academic Preparation, and a Course-Based Intervention. *Journal of Chemical Education* **2018**, <xocs:firstpage xmlns:xocs=""/>.
16.Smith, A. L.; Paddock, J. R.; Vaughan, J. M.; Parkin, D. W., Promoting Nursing Students' Chemistry Success in a Collegiate Active Learning Environment: "if i Have Hope, i Will Try Harder". *Journal of Chemical Education* **2018,** *95* (11), 1929-1938.
17.Talanquer, V., Concept Inventories: Predicting the Wrong Answer May Boost Performance. *Journal of Chemical Education* **2017,** *94* (12), 1805-1810.
18.Casselman, B. L.; Atwood, C. H., Improving General Chemistry Course Performance through Online Homework-Based Metacognitive Training. *Journal of Chemical Education* **2017,** *94* (12), 1811-1821.
19.Cooper, M. M.; Sandi-Urena, S., Design and Validation of an Instrument To Assess Metacognitive Skillfulness in Chemistry Problem Solving. *Journal of Chemical Education* **2009,** *86* (2), 240-245.
20.Herrington, D. G.; Sweeder, R. D., Using Text Messages To Encourage Meaningful Self-Assessment Outside of the Classroom. *Journal of Chemical Education* **2018,** *95* (12), 2148-2154.
21.Mathabathe, K. C.; Potgieter, M., Metacognitive monitoring and learning gain in foundation chemistry. *Chemistry Education Research and Practice* **2014,** *15* (1), 94-104.
22.Parker Siburt, C. J.; MacPhail, R. A.; Bissell, A. N., Developing metacognitive and problem-solving skills through problem manipulation. *Journal of Chemical Education* **2011,** *88* (11), 1489-1495.
23.Parlan, P.; Ibnu, S.; Rahayu, S.; Suharti, S., Effects of the metacognitive learning strategy on the quality of prospective chemistry teacher's scientific explanations. *International Journal of Instruction* **2018,** *11* (4), 673-688.
24.Atkins, P., Chemistry: The great ideas. *Pure and Applied Chemistry* **1999,** *71* (6), 927-929.
25.Cooper, M. M.; Posey, L. A.; Underwood, S. M., Core Ideas and Topics: Building Up or Drilling Down? *Journal of Chemical Education* **2017,** *94* (5), 541-548.
26.Holme, T.; Murphy, K., The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map I: General Chemistry. *Journal of Chemical Education* **2012,** *89* (6), 721-723.
27.Haidar, A. H.; Al Naqabi, A. K., Emiratii high school students' understandings of stoichiometry and the influence of metacognition on their understanding. *Research in Science & Technological Education* **2008,** *26* (2), 215-237.

28.Gulacar, O.; Overton, T. L.; Bowman, C. R.; Fynewever, H., A novel code system for revealing sources of students' difficulties with stoichiometry. *Chemistry Education Research and Practice* **2013,** *14* (4), 507-515.

29.Everson, H. T.; Tobias, S., The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science* **1998,** *26* (1), 65-79.

30.Lindsey, B. A.; Nagel, M. L., Do students know what they know? Exploring the accuracy of students' self-assessments. *Physical Review Special Topics - Physics Education Research* **2015,** *11* (2), 020103.

31.Callender, A. A.; Franco-Watkins, A. M.; Roberts, A. S., Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and Learning* **2016,** *11* (2), 215-235.

32.Zhao, N.; Wardeska, J. G.; McGuire, S. Y.; Cook, E., Metacognition: An Effective Tool to Promote Success in College Science Learning. *Journal of College Science Teaching* **2014,** *43* (4), 48-54.

33.Radovan, M., The Relation between Distance Students' Motivation, Their Use of Learning Strategies, and Academic Success. *Turkish Online Journal of Educational Technology - TOJET* **2011,** *10* (1), 216.

34.Sternberg, R., Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science* **1998,** *26* (1), 127-140.

35.Nelson, T. O.; Dunlosky, J.; Graf, A.; Narens, L., Utilization of Metacognitive Judgments in the Allocation of Study During Multitrial Learning. *Psychological Science* **1994,** *5* (4), 207-213.

36.Miller, T. M.; Geraci, L., Improving metacognitive accuracy: How failing to retrieve practice items reduces overconfidence. *Consciousness and Cognition* **2014,** *29* (C), 131-140.

37.Schraw, G.; Dennison, R. S., Assessing Metacognitive Awareness. *Contemporary Educational Psychology* **1994,** *19* (4), 460-75.

38.Kulak, V. N., Genevieve; Sharma, Rahul, Does the Use of Case-Based Learning Impact the Retention of Key Concepts in Undergraduate Biochemistry? *International Journal of Higher Education* **2017,** *6* (2), 110-120.

39.Horn, S.; Hernick, M., Improving Student Understanding of Lipids Concepts in a Biochemistry Course Using Test-Enhanced Learning. *Chemistry Education Research and Practice* **2015,** *16* (4), 918-928.

40.Lysne, S. J.; Miller, B. G., Research and Teaching: A Comparison of Long-Term Knowledge Retention between Two Teaching Approaches. *Journal of College Science Teaching* **2017,** *46* (6), 100-107.

41.Tucci, V. K.; O'Connor, A. R.; Bradley, L. M., A Three-Year Chemistry Seminar Program Focusing on Career Development Skills. *Journal of Chemical Education* **2014,** *91* (12), 2071-2077.

42.Tamir, P.; Amir, R., High School Science as Viewed by College Students in Israel. *Studies in Educational Evaluation* **1981,** *7* (2), 211-225.

43.Gulacar, O.; Fynewever, H., A Research Methodology for Studying What Makes Some Problems Difficult to Solve. *International Journal of Science Education* **2010,** *32* (16), 2167-2184.

44.Dhlamini, J. J., Enhancing learners' problem solving performance in mathematics: A cognitive load perspective. *European Journal of STEM Education* **2016,** *1* (1), 27-36.

45.Balemans, M. C. M.; Kooloos, J. G. M.; Donders, A. R. T.; Van der Zee, C. E. E. M., Actual Drawing of Histological Images Improves Knowledge Retention. *Anatomical Sciences Education* **2016,** *9* (1), 60-70.

46.Bell, D.; Harless, C.; Higa, J.; Bjork, E.; Bjork, R.; Bazargan, M.; Mangione, C., Knowledge Retention after an Online Tutorial: A Randomized Educational Experiment among Resident Physicians. *Journal of General Internal Medicine* **2008,** *23* (8), 1164-1171.

47.Atmatzidou, S.; Demetriadis, S.; Nika, P., How Does the Degree of Guidance Support Students' Metacognitive and Problem Solving Skills in Educational Robotics? *Journal of Science Education and Technology* **2018,** *27* (1), 70-85.

48.Taasoobshirazi, G.; Glynn, S. M., College students solving chemistry problems: A theoretical model of expertise. *Journal of Research in Science Teaching* **2009,** *46* (10), 1070-1089.

49.Lindsey, R. V.; Shroyer, J. D.; Pashler, H.; Mozer, M. C., Improving Students' Long-Term Knowledge Retention Through Personalized Review. *Psychological Science* **2014,** *25* (3), 639-647.

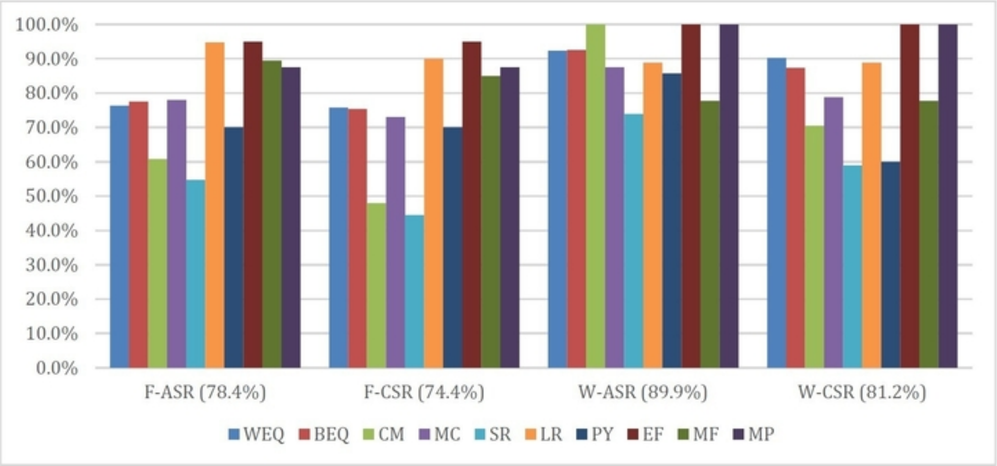


Figure 1. The ASR and CSR scores determined for each subproblem appeared in the problems completed by the focus group students in Fall and Winter quarter.

122x57mm (150 x 150 DPI)

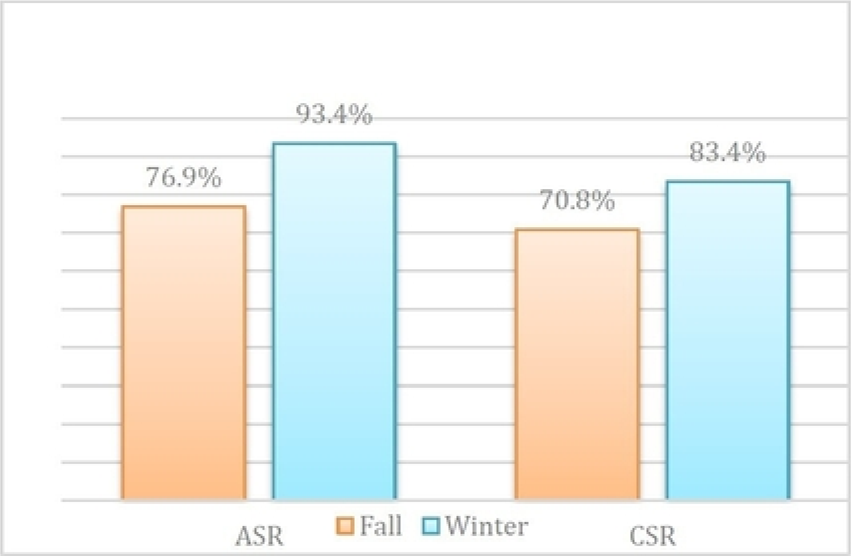
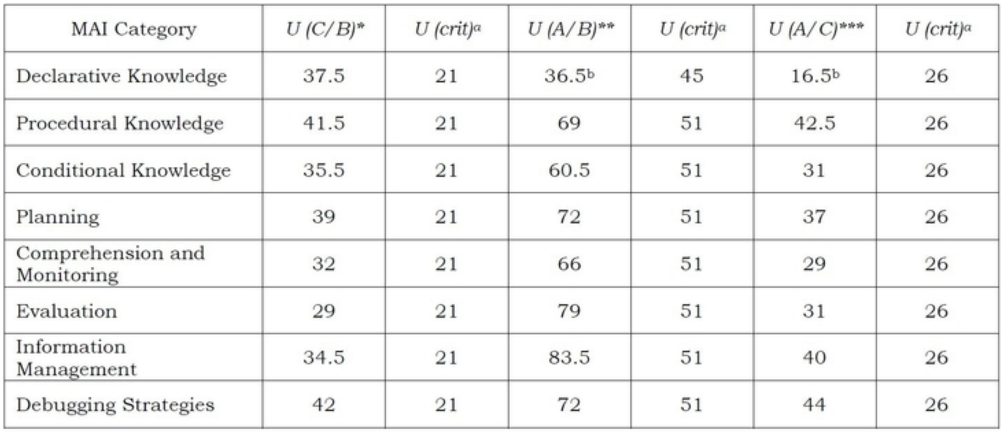


Figure 2. The combined ASR and CSR scores of the focus group for Fall and Winter Quarters

72x47mm (150 x 150 DPI)

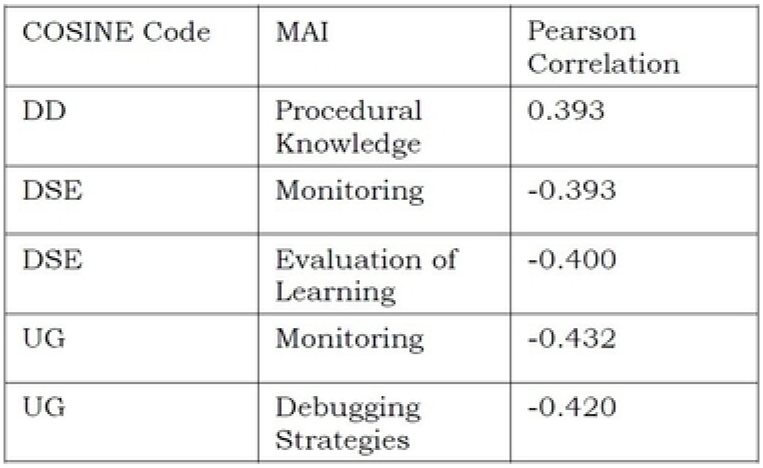
Categories	Codes	Explanation	Example
Successful	Successful (S)	Assigned when student performs subproblem correctly	Student correctly converts grams of a substance into moles of a substance
Neutral	Not Required (NR)	Assigned when student skips a subproblem but uses a different correct method that does not need that subproblem	Student does not find limiting reactant but calculates the number of grams made for each reactant (pretending the rest are in excess), then chooses the smallest gram number
	Did not Know to Do (DD)	Assigned when student skips over a necessary subproblem	Student does not balance chemical equation before doing calculations
	Did Something Else (DSE)	Assigned when student does something different instead of the necessary subproblem	Student adds moles of reactants together instead of finding limiting reactant
Unsuccessful	Unsuccessful-Did Incorrectly (UDI)	Assigned when student understands the needed step but performs it incorrectly	Student makes a math error when converting grams to moles
	Unsuccessful-Received Hint (URH)	Rarely assigned; given when focus group student receives hint from research assistant	Student does not know where to start, so research assistant suggests a step
	Unsuccessful-Guessed (UG)	Assigned when student guesses or it appears that they have guessed	Student writes down an answer with no work preceding it
	Could Not Do (CD)	Assigned when focus group student articulates step needed but cannot do it; also used in the general group for sub steps that students ran out of time to do	Student states that grams must be converted into moles but does not remember to divide by the molar mass

85x67mm (150 x 150 DPI)



MAI Category	U (C/B)*	U (crit)[a]	U (A/B)**	U (crit)[a]	U (A/C)***	U (crit)[a]
Declarative Knowledge	37.5	21	36.5[b]	45	16.5[b]	26
Procedural Knowledge	41.5	21	69	51	42.5	26
Conditional Knowledge	35.5	21	60.5	51	31	26
Planning	39	21	72	51	37	26
Comprehension and Monitoring	32	21	66	51	29	26
Evaluation	29	21	79	51	31	26
Information Management	34.5	21	83.5	51	40	26
Debugging Strategies	42	21	72	51	44	26

123x53mm (150 x 150 DPI)



COSINE Code	MAI	Pearson Correlation
DD	Procedural Knowledge	0.393
DSE	Monitoring	-0.393
DSE	Evaluation of Learning	-0.400
UG	Monitoring	-0.432
UG	Debugging Strategies	-0.420

64x38mm (150 x 150 DPI)