Journal of Humanistic Mathematics

Volume 10 | Issue 2 July 2020

Does Your Course Effectively Promote Creativity? Introducing the Mathematical Problem Solving Creativity Rubric

Kayla K. Blyman
Saint Martin's University

Kristin M. Arney
United States Military Academy at West Point

Bryan Adams
United States Military Academy at West Point

Tara A. Hudson

Recommended Citation

Blyman, K. K. Arney, K. M. Adams, B. and Hudson, T. A. "Does Your Course Effectively Promote Creativity? Introducing the Mathematical Problem Solving Creativity Rubric," *Journal of Humanistic Mathematics*, Volume 10 Issue 2 (July 2020), pages 157-193. DOI: 10.5642/jhummath.202002.09 . Available at: https://scholarship.claremont.edu/jhm/vol10/iss2/9

JHM is an open access bi-annual journal sponsored by the Claremont Center for the Mathematical Sciences and published by the Claremont Colleges Library | ISSN 2159-8118 | http://scholarship.claremont.edu/jhm/

Does Your Course Effectively Promote Creativity? Introducing the Mathematical Problem Solving Creativity Rubric

Cover Page Footnote

All of the authors served on the faculty of the United States Military Academy at West Point in the Department of Mathematical Sciences during varying windows from the Spring 2017 semester through the Spring 2020 semester. The authors would like to thank the Department of Mathematical Sciences at the United States Military Academy at West Point for their support of these efforts.

Does Your Course Effectively Promote Creativity? Introducing the Mathematical Problem Solving Creativity Rubric

Kayla K. Blyman

Department of Mathematics, Saint Martin's University
profkaylablyman@gmail.com

Kristin M. Arney

Department of Mathematical Sciences, United States Military Academy at West Point
kristin.arney.mil@mail.mil

Bryan E. Adams

Department of Mathematical Sciences, United States Military Academy at West Point

Tara A. Hudson

Abstract

As believers in the power of blending the creative with the quantitative, we design our courses with an eye towards developing creative problem solvers. However, when it comes time to evaluate our course's success in developing creative problem solvers we come away with a plethora of qualitative evidence and yet we are left hungry for the quantitative evidence we desire as mathematicians.

In this article we describe the development of the Mathematical Problem Solving Creativity Rubric and its pilot use in a freshman-level Mathematical Modeling and Introduction to Calculus course at the United States Military Academy. We not only come away with the necessary quantitative evidence to satiate our hunger for now, but with a rubric that will allow us to do so in future semesters and courses.

1. Introduction

As mathematics instructors, we hope that our students not only learn the course concepts, but can also apply these ideas in innovative and thoughtful ways. In particular, we hope to foster the ability to solve problems creatively. We go so far as to make developing creative problem solvers one of the higher-order learning goals of our course. However, it is a goal which is often stated and rarely, if ever, evaluated. If creative thought is a course objective, we should be able to monitor student growth in this ability in order to evaluate our successes and failures in addressing this higher-order learning goal. While we acknowledge that assessing creativity is not easy, we believe it is a worthwhile task. Consequently, our goal is to develop and implement a rubric to assess creativity in problem solving primarily for use in course evaluation. In particular, we want to know if there is evidence to support our hypothesis that our Mathematical Modeling and Introduction to Calculus course at the United States Military Academy is improving our students' abilities to solve problems creatively.

To do so, we first consider what characterizes creative thought, and what it means to be creative in a mathematics class in Section 2. Our 5-Point Mathematical Problem Solving Creativity Rubric is presented in Section 3. In Section 4 the methodology of the pilot study using the rubric is discussed. In Section 5 we present the results of the pilot study. Lessons learned from the pilot study and future improvements are discussed in Section 6. Finally, some concluding thoughts are presented in Section 7.

2. Literature Review

J. P. Guilford was among the first to develop a definition of creativity. In 1950 he generated a list of traits that he believed contribute to creative thought [1]. Upon further research, Wilson, Guilford, Christensen, and Lewis identified fourteen factors that contribute to creativity in the sciences including: visualization, ideational fluency, originality, redefinition, adaptive flexibility and spontaneous flexibility [2]. In 1965, Guildford addressed creativity and the implications to high school education. In particular, he summarizes the abilities that contribute to creative thought; these include divergent production abilities (such as fluency, flexibility and elaboration) as well as the ability to transform information to new forms, and ultimately evaluate success [3]. R. J. Sternberg and T. I. Lubart have also conducted research to better understand creativity. According to their investment theory of creativity, the influences on creativity include intellectual abilities, knowledge, styles of thinking, personality, motivation, and environment [4]. In particular, Sternberg and Lubart find that the "willingness to take sensible risks" is a personality trait conducive to creative thought [4, p. 684]. A definition of creativity should be consistent with the work done by these researchers. One such definition is provided by the Association of American Colleges and Universities (AAC&U), they define creativity as: "both the capacity to combine or synthesize existing ideas, images or expertise in original ways and the experience of thinking, reacting and working in an imaginative way characterized by a high degree of innovation, divergent thinking, and risk taking" [5]. The AAC&U was particularly influential to our definition and assessment of creativity.

Since our goal is to assess student creativity in mathematical problem solving, we consider how creativity is defined in a mathematics class. Sriraman discusses mathematical creativity, providing two definitions, one for professional mathematicians and another for K-12 students [6]. He defines creativity in a K-12 mathematics class as: "(a) the process that results in unusual (novel) and/or insightful solution(s) to a given problem or analogous problems, and/or (b) the formulation of new questions and/or possibilities that allow an old problem to be regarded from a new angle requiring imagination" [6, p. 24]. Our population is first-year undergraduate students with a wide variety of interests and talents; it is likely that only a small fraction will become mathematics majors. Therefore, when assessing creativity in mathematical problem solving for this population, it is appropriate to adopt the mindset that Sriraman gives for mathematical creativity in high schools. After synthesizing the research, we formulate the following definition of creativity in mathematical problem solving: students can extend knowledge to new situations, draw upon previous experiences, develop illustrations to clarify concepts, establish connections between concepts, and take responsible risks.

Amabile provides guidance on assessing creativity based on her consensual definition of creativity [7]. According to her definition, "...creativity can be regarded as the quality of products or responses judged to be creative by appropriate observers, and it can also be regarded as the process by which something so judged is produced." [7, p. 1001]. Therefore, the creativity of a product should be assessed by an expert in the field. During implementation, Amabile suggests assessing the level of creativity in comparison to similar samples [7]. Several rubrics have been created with the goal of assessing the creativity of a process or a product [8, 9, 5, 10, 11, 12].

While all of the facets of our definition of creativity in mathematical problem solving are process oriented, we develop our rubric to assess evidence of these creative processes within the product. This idea is not unprecedented. Young creates a rubric to assess creativity in writing by looking for evidence of creative thought [12]. When creating our rubric to assess creativity in problem solving, we adopt a similar approach. One creativity rubric that was particularly influential on our rubric as well as on the academic program goals at the United States Military Academy was the Creative Thinking VALUE Rubric created by the AAC&U [5]. The facets we chose to include in our rubric are: *Originality*, *Flexibility*, *Visualization*, *Elaboration*, and *Risk*.

- *Originality* refers to the student's ability to extend knowledge to new situations. Originality has been linked to creativity by researchers such as Wilson et al., and Sternberg and Lubart [2, 4]. Further, the AAC&U includes this facet in their Creative Thinking Value Rubric as "Innovative Thinking" [5]. To allow originality to be assessed relative to similar products, a random sample of papers were evaluated independently, the originality scores were discussed, and then this facet was calibrated. This follows the suggestion given by Amabile in [7].

- *Flexibility* is the ability to integrate knowledge and skills from a variety of disciplines. Wilson et al. establish the connection between creativity and flexibility in their discussion of adaptive flexibility and spontaneous flexibility [2].

- *Visualization* is the act of developing illustrations to clarify concepts. While it is certainly possible that a student is able to arrive at a creative solution without a picture, providing an illustration can be evidence that creative thought has taken place. Visualization is a facet found to contribute to creative thinking according to Wilson et al. [2]. Furthermore, the process of creating an illustration involves transforming the problem which is included in the AAC&U Creative Thinking Value Rubric [5].

- *Elaboration* is the ability to establish meaningful connections between concepts and to explain a thought process in words. As with visualization, elaboration is not essential for a creative solution, but it is potential evidence of creative thought. It is a way for the evaluator to see the student's thought process, and the connections that they have made. This facet is included in the AAC&U Creative Thinking Value Rubric in the "Connecting, Synthesizing Transforming" component [5].
- *Risk* refers to taking responsible risk in the problem solving process. Sternberg and Lubart tie taking reasonable risks to creativity, as summarized in [4]. Further, taking risks is also a component of the AAC&U Creative Thinking Value Rubric [5].

3. The 5-Point Rubric

After consulting the literature and choosing the five facets for our rubric based on the work of others, we set out to develop a rubric that was clear, concise, broadly applicable to mathematical problem solving, and ultimately easy to use. While the AAC&U provides a well-known rubric for Creative Thinking [5], their rubric is interdisciplinary and does not address the many ways students may show creativity in mathematical problem solving when working on problems with definite known answers. It is our belief that students can show creativity when approached with a mathematics problem that is not an open problem. Therefore, we sought to embrace the many ways creativity may appear in all levels of mathematical problem solving in our rubric. After several previous versions, we arrived at the rubric found in Figure 1.

Because the rubric is broadly applicable in mathematical problem solving, it is important that anyone using the rubric think about the problem that they are planning to evaluate and consider what specifically some of the entries in the rubric may mean for them. For example, in one problem using a volume formula may be at least moderately original, while in another problem it may simply be a standard formula. Additionally, we encourage anyone evaluating originality to make sure they take the time to calibrate the rubric to the student responses received for the problem. This is discussed further in Section 6.

Facet	**Description**	**Score: 1**	**Score: 2**	**Score: 3**	**Score: 4**	**Score: 5**
Originality	*Extends Knowledge to New Situations*	Standard Formulas or Brute Force	Some Potential for Some Abstraction (Variables or Assumptions)	Low Abstraction: Previously Taught Techniques	Insightful Observations Used to Solve	High Abstraction: Beyond Taught Techniques
Flexibility	*Uses Multiple or Interdisciplinary Approaches*	Partial Attempt	Approaching a Solid Attempt (i.e. All But Solution)	Solid Attempt: Results in a Conclusion	Possible Other Attempt or Same Method Multiple Ways	More Than One Attempt
Visualization	*Develops and Uses Illustrations to Clarify Concepts*	None (or No Apparent Use of Provided Graphic)	Minimal	Something but no obvious connection or use	Possible Connection, but not clear or complete	Picture with Connection to Methodology
Elaboration	*Establishes Meaningful Conceptual Connections*	None	Minimal (i.e. Labels)	Some attempt at making a meaningful connection	Comprehensible attempt but lacking clarity or detail	Interpretation of Result Providing True Connection
Risk	*Takes Responsible Risks*	Blank	Not much work and no answer	Not Much Work with an Answer	Lots of work and no answer	Lots of work with an answer

Figure 1: 5-Point Mathematical Problem Solving Creativity Rubric

4. Methods

Our primary goal in creating this rubric is to evaluate how our course is doing at achieving our higher-order learning goal of developing creative problem solvers. During the Fall 2018 semester we ran a pilot study to help us develop the rubric and put it to the test. The pilot study was conducted in Mathematical Modeling and Introduction to Calculus at the United States Military Academy. This course is taken by about 75% of the incoming first-year students during their first fall semester. The course first introduces students to modeling using discrete dynamical systems consisting of a single recursion equation, followed by modeling with systems of recursion equations, and finally modeling with continuous functions as a preparation for their subsequent calculus course. Like many other mathematics courses, one of the higher-order learning goals of this course is to develop creative problem solving abilities. We developed this rubric and pilot study with the goal of being able to evaluate how well we were doing in achieving this higher-order learning goal as a course.

4.1. Sampling

In the Fall 2018 semester, 835 students were enrolled in Mathematical Modeling and Introduction to Calculus. The 835 students were split into 50 sections that were taught by a total of 19 instructors. 10 of the 19 instructors had taught college mathematics at the United States Military Academy or another institution the previous year. All of these instructors were asked to participate with all sections that they were teaching. Of those 10 instructors, 8 (80%) opted in. This created a sample for the pilot study consisting of 318 out of 835 (38%) students, in 19 out of 50 (38%) sections offered which were taught by 8 out of 19 (42%) instructors overall for the course. Additionally, the course was offered during four different hours throughout the day and was sectioned into three different types of sections based on performance on an assessment of fundamental mathematics concepts. The 19 participating sections were spread relatively evenly across the hours and across performance levels on the fundamental mathematics concepts assessment. These sections were split into a group of 10 sections (164 students) and a group of 9 sections (153 students). Group 1 had 164 students complete the pre-test and 149 students complete the post-test. Group 2 had 153 students complete the pre-test and 148 students complete the post-test. Each group represented a cross section of instructors, hours the course was offered, and performance on the fundamental mathematical concepts assessment.

4.2. Study Design

Since no prior studies of this kind that we are aware of have looked at creativity in mathematical problem solving, there are no problems with benchmarks for student performance with regards to creativity in their problem solving. Consequently, we had approximately half of the sample for the study complete Problem A as a pre-test and Problem B as the post-test while the other half of the study population completed Problem B as a pre-test and Problem A as the post-test in an attempt to control for the natural differences that may arise between problems. Which treatment each group received was randomly assigned between the two groups. This resulted in 10 sections completing the cheese problem presented below as the pre-test (Group 1) and 9 sections completing the surface area problem presented below as the pre-test (Group 2). The pre-test was administered during the first week of classes and the post-test was administered during the last week of classes, 17 weeks later.

4.3. The Problems

Much careful thought was put into the selection of problems. It was important that the problems were able to be solved given the expected content knowledge of our students, but it was equally important that course content not be helpful in solving either of the problems so that there was no content advantage to taking either problem as a post-test rather than a pre-test. Additionally, the problems chosen needed to be unique enough that the students would feel like they had never seen the problem before for both the pre-test and the post-test regardless of the order in which they received the problems. Finally, the problems needed to not only be able to be explored and ultimately solved in a variety of ways, but to be difficult enough that the students would truly be challenged to solve them in any way.

The first problem selected was the cheese problem from an author's graduate course work:

> A $3 \times 3 \times 3$ cube of cheese is divided into 27 small ($1 \times 1 \times 1$) cubes. A mouse eats one small cube each day and an *adjacent*

> small cube (that is sharing a face) the next day. Can the mouse eat the *center* small cube on the last day?

The second problem selected was the surface area problem from the 2011 Mathematical Festival [13, p. 2]:

> A wooden block is divided into eight smaller blocks by three cuts. In the figure, the surface areas of the seven visible blocks are labeled. What is the surface area of the eighth block?

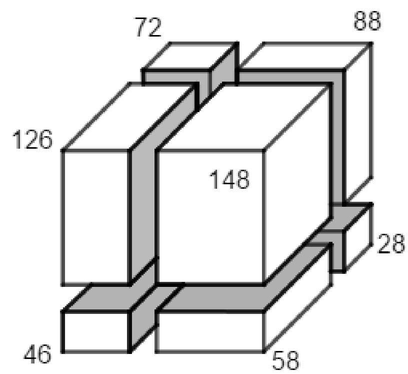


Figure 2: Surface Area Problem Provided Figure

For each problem students were given ten minutes for completion without technology and the following directions.

> Do your best to solve the following problem. Show **ALL** of your work. Explain your reasoning. Do not erase or cross out anything.

4.4. Data Analysis

When scoring the problems, pre-test and post-test problems were all put together and the difference between them was unidentifiable to the two raters. Each problem had a student's identifying number on it to allow it to be paired for a pre-test to post-test comparison and to know which problem was the pre-test and post-test. The two raters were the developers of the rubric and each scored all problems. Common themes and extremely unique examples of student work were regularly discussed as we worked to calibrate the rubric. Across 3,075 individual scores given on a 5-point scale, scores between the two raters differed by more than one point 11.12% of the time.

Moreover, of the 12,300 possible points of separation between the scores given by the raters, only 13.38% of that separation occurred. In the end, the scores used for analysis were the average of the two raters whenever there was disagreement.

The main question that drove the study is "Are we succeeding in achieving our higher-order learning goal of developing creative problem solvers?" To address this question we conducted a paired t-test to assess if, on average, the students performed better on their post-test problem. Additionally, we conducted paired t-tests to assess if students on average improved on their post-test problem in both groups. Lastly, we conducted two-sample t-tests to assess how the pre-test scores, on average, compared to the post-test scores for each problem. In all t-tests, the difference in scores were determined by (post-test score) − (pre-test score). The null hypothesis for each t-test was that there was no difference between the pre-test and post-test scores. The alternative hypothesis was that there is a difference between the pre-test and post-test scores (two-sided). R was utilized to perform all paired t-tests[14]. T-tests were utilized because the pre-test and post-test scores approximately followed a normal distribution as well as their paired differences.

5. Results

Summary statistics of the scores for the cheese problem, Table 1, and surface area problem, Table 2, were calculated using the pre-test and post-test scores to assess how the different natures of the two problems might have influenced the scores for each facet. On average, the cheese problem scores were higher than the surface area problem scores. This could imply that the natures of the problems made students less likely to use creative problem-solving methods on the surface area problem than on the cheese problem.

5.1. Comparing Pre-test to Post-test Scores

Results from the paired t-test comparing the pre-test and post-test scores are found in Table 3. The average difference was calculated using (post-test score)− (pre-test score). A positive value implies, on average, the students improved between the pre-test and post-test. The mean difference was positive for each facet except visualization and elaboration. Flexibility (p-value = 0.019) and

Facet	Mean	Minimum	Q1	Median	Q3	Maximum
Originality	2.72	1.00	2.00	3.00	3.50	5.00
Flexibility	3.13	1.00	3.00	3.00	3.50	5.00
Visualization	2.95	1.00	2.50	3.00	3.50	5.00
Elaboration	3.00	1.00	2.00	3.00	4.00	5.00
Risk	3.90	1.00	3.00	4.00	5.00	5.00
Total	15.70	5.00	13.50	16.00	18.50	22.00

Table 1: Summary statistics for the cheese problem

Facet	Mean	Minimum	Q1	Median	Q3	Maximum
Originality	2.52	1.00	1.00	2.50	3.50	5.00
Flexibility	2.90	1.00	1.50	3.00	4.00	5.00
Visualization	2.22	1.00	1.50	2.00	3.00	5.00
Elaboration	2.42	1.00	1.12	2.00	3.50	5.00
Risk	3.30	1.00	2.50	3.00	4.00	5.00
Total	13.37	6.00	10.50	13.50	16.00	22.50

Table 2: Summary statistics for the surface area problem

risk (p-value = 0.019) provided strong evidence against the null hypothesis. For both measures, the 95% confidence interval contains only positive values, demonstrating that on average there was an increase in scores for these facets.

5.2. Comparing Pre-test to Post-test Scores by Group

Results from the paired t-test comparing the pre-test and post-test scores for each group are in Tables 4 and 5. Table 4 contains the results for students in Group 2 that took the surface area pre-test and the cheese post-test. Table 5 contains the results for students in Group 1 that took the cheese pre-test and the surface area post-test.

For Group 2, Table 4, the mean difference was positive for each facet and all t-tests had p-values < 0.01. Each facet's 95% confidence interval contains only positive values showing that on average the students' scores in this group increased between pre-test and post-test.

Facet	Mean Difference	t-statistic	p-value	95% CI
Originality	0.154	1.636	0.103	(-0.031, 0.339)
Flexibility	0.194	2.352	0.019	(0.032, 0.357)
Visualization	-0.079	-1.027	0.305	(-0.231, 0.073)
Elaboration	-0.017	-0.183	0.855	(-0.198, 0.165)
Risk	0.194	2.368	0.019	(0.033, 0.356)
Total	0.446	1.496	0.136	(-0.141, 1.032)

Table 3: Paired t-test results for all students

Facet	Mean Difference	t-statistic	p-value	95% CI
Originality	0.357	2.630	0.009	(0.089, 0.625)
Flexibility	0.441	3.570	< 0.001	(0.197, 0.685)
Visualization	0.696	8.150	< 0.001	(0.527, 0.865)
Elaboration	0.57	4.559	< 0.001	(0.323, 0.817)
Risk	0.801	7.035	< 0.001	(0.576, 1.026)
Total	2.864	7.086	< 0.001	(2.065, 3.663)

Table 4: Paired t-test results for students with the surface area pre-test and the cheese post-test (Group 2)

For Group 1, Table 5, the mean difference was negative for each facet and all t-tests for the facets, except originality and flexibility, had p-values < 0.01. With the exception of originality and flexibility, each facet's 95% confidence interval contains only negative values showing that on average the students' scores in this group decreased between pre-test and post-test.

5.3. Comparing Pre-test to Post-test Scores by Problem

Results from the two-sample t-test comparing the pre-test and post-test scores for each problem are in Tables 6 and 7. Table 6 contains the results for the surface area problem. Table 7 contains the results for the cheese problem.

For the surface area problem, all but elaboration had a positive difference between pre-test and post-test scores. There were statistically significant increases in scores from pre-test to post-test for the surface area problem in the flexibility (p-value = 0.002) and risk facets (p-value < 0.001) for the

Facet	Mean Difference	t-statistic	p-value	95% CI
Originality	-0.036	-0.279	0.781	(-0.290, 0.219)
Flexibility	-0.036	-0.334	0.739	(-0.249, 0.177)
Visualization	-0.804	-8.519	< 0.001	(-0.990, -0.617)
Elaboration	-0.565	-4.747	< 0.001	(-0.801, -0.330)
Risk	-0.373	-3.807	< 0.001	(-0.566, -0.179)
Total	-1.814	-5.207	< 0.001	(-2.502, -1.126)

Table 5: Paired t-test results for students with the cheese pre-test and the surface area post-test (Group 1)

surface area problem. For the cheese problem, only elaboration and risk had a positive difference between pre-test and post-test scores. There were not any statistically significant differences in the pre-test and post-test scores for the cheese problem.

Facet	Difference of Means	t-statistic	p-value	95% CI
Originality	0.39	2.732	0.007	(0.109, 0.671)
Flexibility	0.459	3.168	0.002	(0.174, 0.745)
Visualization	0.03	0.270	0.788	(-0.193, 0.254)
Elaboration	-0.077	-0.561	0.575	(-0.349, 0.194)
Risk	0.429	3.844	< 0.001	(0.209, 0.649)
Total	1.232	2.853	0.005	(0.382, 2.082)

Table 6: Two-sample t-test comparing the surface area problem pre-test and post-test scores

5.4. Discussion

Evaluating creativity is a difficult task. Firstly, it is challenging to assign a number for each of the facets of creativity based on student written work. Appendix A provides an annotated spectrum of the solutions received for each problem. Secondly, it is perhaps more challenging to meaningfully interpret the consequent quantitative results. Overall, we feel that our results may be somewhat confounded due to the nature of the problems chosen for this experiment but we do find promise in the abilities of our students in flexibility and risk when using our rubric.

Facet	Difference of Means	t-statistic	p-value	95% CI
Originality	-0.09	-0.828	0.408	(-0.305, 0.124)
Flexibility	-0.044	-0.435	0.664	(-0.245, 0.156)
Visualization	-0.111	-1.257	0.21	(-0.284, 0.063)
Elaboration	0.061	0.501	0.617	(-0.179, 0.302)
Risk	0.035	0.315	0.753	(-0.182, 0.252)
Total	-0.149	-0.382	0.703	(-0.920, 0.621)

Table 7: Two-sample t-test comparing the cheese problem pre-test and post-test scores

Table 5 tells us that students in Group 1 who completed the cheese problem as the pre-test and the surface area problem as the post-test had a statistically significant decrease in almost every facet of their creative problem solving abilities throughout our course. Alternatively, Table 4 tells us that students in Group 2 had a statistically significant increase in every facet of their creative problem solving abilities throughout our course. Not surprisingly, Table 3 reveals that when we put everyone together we do not see as many statistically significant changes. Ultimately, the results in Table 3 indicate that the increases seen in Table 4 outweigh the decreases seen in Table 5 for the flexibility and risk facets, and for all of the other facets they more or less balance one another.

Considering the summary statistics for the cheese problem presented in Table 1 compared to the summary statistics for the surface area problem presented in Table 2 it is easy to notice that the scores for the cheese problem were higher than the scores for the surface area problem on average. Consequently, it is reasonable to believe that the difference in type of solution required for each problem as well as the difference in tools provided for students within the statement of each problem confounded the results seen in Tables 3, 4, and 5. In fact, considering the differences in means from the cheese problem to the surface area problem and the 95% confidence intervals provided in Tables 4 and 5, it is possible that the difference in means between the problems accounted for all of the differences in means seen from pre-test to post-test with the exception of one. In the case of the risk facet when the cheese problem was the pre-test and the surface area problem was the post-test, the difference in means between the problems leads to the expectation of a -0.6 difference from pre-test to post-test and -0.6 is below the 95% confidence interval, making the decrease in the risk facet from pre-test to post-test less than we might expect.

Given the difference in problem means, it is reasonable to expect the combined results to skew towards a decrease since more students completed the cheese problem as a pre-test. However, Table 3 reveals that there was a statistically significant increase in the flexibility and risk facets which is opposite of what the difference in problem means and the number of students completing each problem would lead us to expect. Therefore, we conclude that the increases in flexibility and risk from pre-test to post-test for our combined sample are truly significant.

While we are encouraged by that positive result, perhaps the more convincing result is found in Table 6. Here we find the results comparing the pre-test scores for the surface area problem to the post-test scores for the same problem. While it is important to note that the students completing the surface area pre-test were not the same as the students completing the surface area post-test, the two groups were assigned using random assignment to ensure as much balance as possible between the two groups. Further, if the two groups were out of balance, we should expect to see evidence of that in Table 7 as well since the group that completed the surface area pre-test completed the cheese post-test and vice versa. However, in Table 6 we see statistically significant increases in originality, flexibility, risk, and in the total of all five facets, while in Table 7 we see relatively neutral results across all facets with no statistical significance. Overall, the cheese problem brought with it a lot of challenges. In Section 6, we will discuss further the difficulties that were realized as we evaluated the student responses to that problem.

6. Lessons Learned and Future Improvements

The pilot use of the Mathematical Problem Solving Creativity Rubric went relatively smoothly and we are encouraged by the results. However, reflecting on the experience reveals a few things that should be carefully considered, and potentially addressed, in future iterations.

Facet	**Description**	**Score: 1**	**Score: 2**	**Score: 3**	**Score: 4**	**Score: 5**
Originality	*Extends Knowledge to New Situations*	Standard Formulas or Brute Force	Some Potential for Some Abstraction (Variables or Assumptions)	Low Abstraction: Previously Taught Techniques	Insightful Observations Used to Solve	High Abstraction: Beyond Taught Techniques
Flexibility	*Uses Multiple or Interdisciplinary Approaches*	Partial Attempt	Approaching a Solid Attempt (i.e. All But Solution)	Solid Attempt: Results in a Conclusion	Possible Other Attempt or Same Method Multiple Ways	More Than One Attempt
Risk	*Takes Responsible Risks*	Blank	Not much work and no answer	Not Much Work with an Answer	Lots of work and no answer	Lots of work with an answer

Figure 3: Reduced 5-Point Mathematical Problem Solving Creativity Rubric

6.1. The Rubric

When initially selecting facets for the rubric, we believed it would be helpful to evaluate visualization and elaboration as both are identified as being potential evidence of creative problem solving. However, upon evaluating the problems in our pilot study and considering the quantitative results, we believe that evaluating these two facets was largely unhelpful in determining a student's creative problem-solving ability. Rather, we found ourselves relying on what the students communicated to us through visual and elaborative means to help us determine the other three facets of creativity. The literature warned that this may be the case as it identifies both visualization and elaboration as potential evidence for creativity rather than necessary facets for creative thought [1, 3, 2]. Combining our experiences with the literature, we have come to the conclusion that producing a separate score for visualization and elaboration is unnecessary and have consequently made the rather significant decision to simplify our rubric to include only three facets – originality, flexibility, and risk – for future uses. Figure 3 provides the reduced 5-point rubric.

6.2. Problem Selection

When selecting problems for the pilot, the cheese problem and the surface area problem seemed relatively similar in that they both did not have an obvious, straight-forward, solution and they could both be solved in multiple ways. However, as we evaluated the two problems, the differences between them became apparent. While we were initially excited about the creative possibilities the cheese problem brought with it, we ultimately became frustrated by it.

One phenomenon that we observed was the seeming desire of students to do whatever they could to manipulate the situation to make it possible for the answer to the question to be "yes." While our student population may be particularly predisposed to make it work one way or another, it seems plausible that any student population may be inclined to respond in the affirmative to a problem posed by their instructor. Consequently, we recommend avoiding Yes/No problems to avoid this potential influence on student responses.

Another potential issue identified with the cheese problem was the large amount of work that students could do on the problem through thought experiments without ever writing anything down. Despite the explicit instructions to write everything down and not erase or cross out anything, it was apparent that there were unrecorded thoughts, erased work, and crossed out work on several of the submissions. While it was not as prevalent with the surface area problem, the same phenomenon occurred on a few of those submissions as well. While some of this can be addressed by selecting problems that make thought experiments without writing difficult, this consideration may not be enough. Evidence for this is seen in the few surface area problems which had a similar phenomenon occur, despite the problem's potential to be a more computational problem. Seeing all of these responses that clearly had more to them than what was recorded leads to the desire to gain deeper insight into what students are thinking and not recording as they work towards solving these problems. It seems the best way to do this moving forward would be to have some students complete the problems in an interview setting where they can both write and speak their problem-solving thoughts and an interviewer can ask questions to prompt them to share their thoughts if they are not naturally doing so. Not only would this allow for deeper insight into the interviewed students creative problem-solving thoughts, but it would allow a comparison to be made between the totality of the students' communicated thoughts and what made its way onto paper.

One way that the two problems varied greatly was in how much insight we could gain as evaluators into the students' thoughts when there was little to no elaboration present. Without students telling us their thought processes we could often figure out what they were trying to do, and consequently what they were likely thinking, on the surface area problem. However, when there was little to no elaboration on the cheese problem it was almost always impossible to tell what the thought process might have been. While there is always going to be some amount of guessing involved when students fail to elaborate well, some problems will require more than others. Consequently, it is important to consider how much guessing may be involved when selecting problems.

One of the most frustrating issues that we encountered was students misinterpreting and reinterpreting the problems posed in ways that made the problems completely void of any richness. In these reinterpretations, the answers should be things like "yes, of course the mouse can eat the center cube of cheese eventually," or "yes, if the mouse eats 27 cubes of cheese in one day the mouse will eat the center cube on the last day." Not only is there no need for creativity in answering these questions there is no real need for problem solving either. Whether these misinterpretations and reinterpretations occurred due to reading comprehension issues or due to uncertainty about how to solve the problem as originally posed, they raised issues in evaluation and should consequently be considered when selecting problems. Again while all problems can be misinterpreted or reinterpreted, some are more likely to be misinterpreted or reinterpreted. In the case of the pilot, the cheese problem was much more regularly misinterpreted and reinterpreted compared to the surface area problem.

6.3. Using the Rubric

If it is to be relevant, the rubric needs to remain useful when students solve problems in unexpected ways. The Mathematical Problem Solving Creativity Rubric did a pretty good job at allowing us to adapt to all kinds of circumstances. If things do not go as planned, as with misinterpretations and reinterpretations of the problem being prevalent, it is important to decide in the beginning how to handle such occurrences and to remain consistent in your treatment of them.

While almost all of the facets on the rubric do not require any real calibration, the originality facet does. For the pilot study we looked at roughly 10% of the submissions for each problem before deciding what problem-solving processes were placed in each of the categories. After scoring all of the submissions, we realized a need for a larger sample size when calibrating that facet. In the future we will attempt to calibrate it using a sample size of approximately 25% of the submissions.

There were several occurrences of students turning the problem into a more elaborate short story, making seemingly random and unnecessary assumptions about the problem, or drawing the whole situation out as a cartoon. While all of these are most certainly evidence of creativity, they are not necessarily evidence of creative problem solving. When using the rubric, it can be tempting to get drawn into these displays of creativity sans problem solving, but it is important to remember that we are ultimately evaluating creative problem-solving ability, not simply creativity. As you evaluate the papers it is advantageous to do periodic sanity checks that sound something like “is this creative problem solving or just creative?”

It is not uncommon when evaluating these submissions to find yourself on a roller coaster of gut feelings about the submissions and questioning your consistency. Creative problem solving is very difficult to evaluate. While we believe the rubric is good because of how consistent we were between two evaluators over all of the submissions, we both had our moments of feeling like everything was awful. Because creative problem solving is so nebulous, we recommend maintaining more than one evaluator per problem if at all possible. If that is not possible, then having a single instructor evaluate all submissions and then scramble them up and blindly evaluate them a second time would suffice. This practice allows the evaluators to help keep each other honest and balanced with scores during the high and low times through comparison and ultimately through averaging the scores.

6.4. Using the Scores

No matter how consistent the scoring is, no two problems are going to look alike on all facets of creative problem solving. Because of this, the split sample with two problems is desirable so that overall summary statistics including both pre-test and post-test submissions for each problem can be established to help make the interpretation of the results from pre-test to post-test meaningfully possible. If you are using a problem and the rubric to get an idea of where your students are in their creative problem-solving abilities, it may be desirable to use a problem that you know a pre-test and post-test overall average for, or to at least think about how the type of solution required as well as the tools provided within the statement of the problem may effect the student responses in each of the facets.

6.5. Potential 3-Point Rubric

Along the way to developing our 5-Point Rubric, we developed a simplified 3-Point version of the rubric. Unfortunately, this version of the rubric did not differentiate enough within each facet to be useful for our desired result of evaluating our course's successes or failures at achieving our higher-order learning goal of developing creative problem solvers. However, we kept this version of the rubric because it was incredibly fast and easy to use in comparison to the 5-Point version.

Consequently, depending on your intended use of the rubric, the 3-Point version may be preferable. For example, sacrificing some differentiation within facets for speed may be desirable if the rubric is being used to gain insight into the current creative problem-solving abilities of a class. This insight may be useful in determining how much scaffolding to provide a class to most effectively develop creative problem solvers.

One challenge that may arise when using the rubric in this way are the broad range of solutions that may get lumped into the Score: 2 category if Score: 1 is allowed to mean awful and Score: 3 is allowed to mean amazing. By allowing student submissions that are not amazing but are pretty good to be placed in the Score: 3 category and submissions that are not awful but are pretty weak to be placed in the Score: 1 category, the Score: 2 category will retain some meaning. Another challenge that may arise is figuring out how to assign any meaning to the scores on student submissions without a comparison method.

7. Conclusion

Despite the many challenges involved in assessing creative problem solving, it is a worthwhile pursuit. As we seek to foster students' creativity in our classrooms, we include higher-order learning goals to develop creative problem solvers in our course goals because we know that the future will need creative problem solvers. However, without the ability to evaluate our successes and failures at fostering student creativity, we run the risk of maintaining a status quo that falls short of truly developing creative problem solvers. If we fail – even unknowingly – we will be leaving the future without the ingenuity its problems demand. Therefore, no discussion of fostering creativity in mathematics classrooms is complete without addressing the need for evidence of such.

This research resulted in the creation of a rubric that assists us in assessing our course design as we strive to develop creative problem solvers through experiences in our course. Based on our initial findings, we know that our course is successfully addressing our higher-order learning goal of preparing our students to be problem solvers, leaders, and decision makers in a world that will demand creative problem solving.

The Mathematical Problem Solving Creativity Rubric makes the process of evaluating creative problem solving much more approachable. We will be using it in the future and we hope that you will join us in evaluating how your own courses are doing in their efforts to develop creative problem solvers.

Acknowledgements

All of the authors served on the faculty of the United States Military Academy at West Point in the Department of Mathematical Sciences during varying windows from the Spring 2017 semester through the Spring 2020 semester. The authors would like to thank the Department of Mathematical Sciences at the United States Military Academy at West Point for their support of these efforts.

References

[1] J. P. Guilford. Creativity. *American Psychologist*, 5(9):444 – 454, 1950. ISSN 0003-066X. URL https://usmalibrary.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=1951-04354-001&site=ehost-live&scope=site.

[2] Robert Wilson, J. Guilford, Paul Christensen, and Donald Lewis. A factor-analytic study of creative-thinking abilities. *Psychometrika*, 19(4):297, 1954. ISSN 00333123. URL https://usmalibrary.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=71554660&site=eds-live&scope=site.

[3] J. P. Guilford. Creativity in the secondary school. *The High School Journal*, 48(8):451–458, 1965. ISSN 00181498, 15345157. URL http://www.jstor.org/stable/40366066.

[4] Robert J. Sternberg and Todd I. Lubart. Investing in creativity. *American Psychologist*, 51(7):677 – 688, 1996. ISSN 0003-066X. URL https://usmalibrary.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=1996-04968-001&site=ehost-live&scope=site.

[5] Association of American Colleges and Universities (AAC&U). Creative thinking VALUE rubric, 2009. URL https://www.aacu.org/value/rubrics/creative-thinking.

[6] Bharath Sriraman. Are giftedness and creativity synonyms in mathematics?. *Journal of Secondary Gifted Education*, 17(1):20 – 36, 2005. ISSN 1077-4610. URL https://usmalibrary.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ746043&site=eds-live&scope=site.

[7] Teresa M. Amabile. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43(5):997 – 1013, 1982. ISSN 00223514. URL https://usmalibrary.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=16645646&site=ehost-live&scope=site.

[8] Houssein El Turkey, Gail Tang, Milos Savic, Gulden Karakok, Emily Cilli-Turner, and David Plaxco. The creativity-in-progress rubric on proving: Two teaching implementations and students' reported usage. *PRIMUS*, 28(1):57 – 79, 2018. ISSN 1051-1970. URL https://usmalibrary.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1171844&site=ehost-live&scope=site.

[9] Renee M. Clary, Robert Brzuszek, and C Fulford. Measuring creativity: A case study probing rubric effectiveness for evaluation of project-based learning solutions. *Creative Education*, 2:333–340, 01 2011. doi: 10.4236/ce.2011.24047.

[10] Milos Savic, Gulden Karakok, Gail Tang, Houssein El Turkey, Emily cilli Turner, Kimberley Cadogan, and Paul Regier. Creativity in progress rubric, 2018. URL http://www.creativityresearchgroup.com/.

[11] Kate Shively, Krista M. Stith, and Lisa DaVia Rubenstein. Measuring what matters: Assessing creativity, critical thinking, and the design process. *Gifted Child Today*, 41(3):149 – 158, 2018. ISSN 1076-2175. URL https://usmalibrary.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1183587&site=ehost-live&scope=site.

[12] Linda Payne Young. Imagine creating rubrics that develop creativity. *English Journal*, 99(2):74 – 79, 2009. ISSN 0013-8274. URL https://usmalibrary.idm.oclc.org/login?url=https:

//search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ863954&site=ehost-live&scope=site.

[13] Ivan Yashchenko. *Invitation to a Mathematical Festival*. MSRI and AMS, 2013.

[14] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL https://www.R-project.org/.

A. Sample Student Work

Responses to each problem varied widely. What is presented here represents only a selection of the problem solving approaches and methods of communication received.

Figure 4 is an example of a brute force approach to solving the surface area problem. In this example, the student has added the seven provided surface areas of the smaller blocks along with a few sums of subsets of these surface areas. Ultimately they did not provide an answer to the question posed. This submission received an originality score of 1, a flexibility score of 1, a visualization score of 1, an elaboration score of 1, and a risk score of 2.

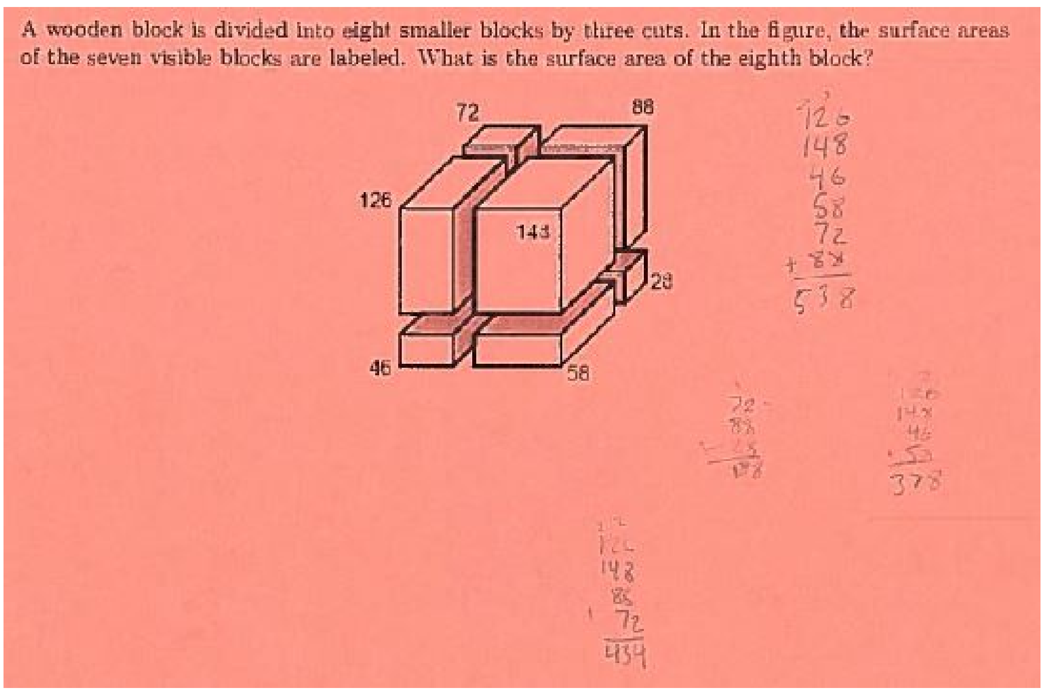


Figure 4: A brute-force attempt of the surface area problem.

Figure 5 is an example of an approach to solving the surface area problem where the student did not think beyond surface area formulas. In this example, the student has added the seven provided surface areas of the smaller blocks and has provided the surface area formula as well as dividing some of the provided surface areas by two to determine the outer surface area of the

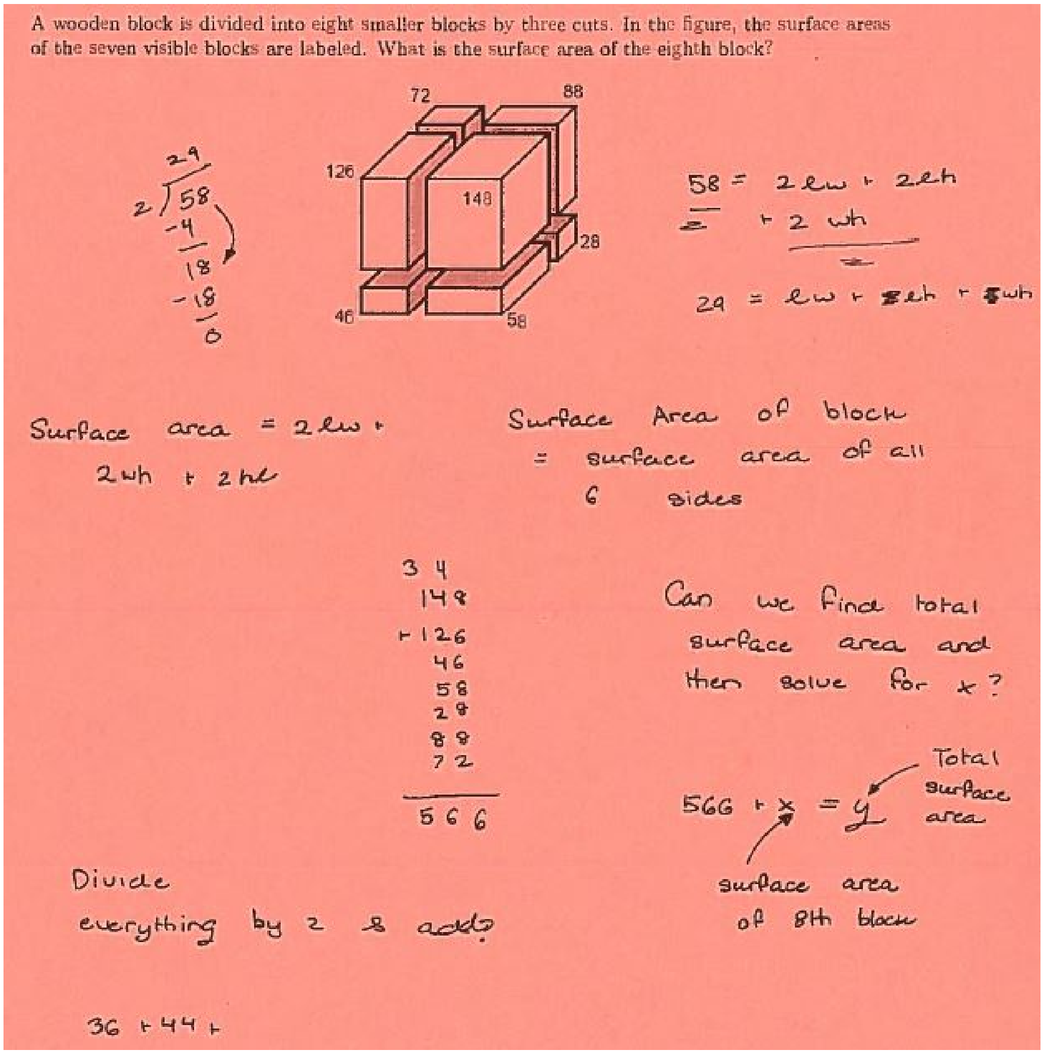


Figure 5: An attempt at the surface area problem using only surface area.

larger block. Although they provided an equation which included a variable which was defined to be the solution, ultimately they did not provide an answer to the question posed. This submission received an originality score of 1, a flexibility score of 2, a visualization score of 1, an elaboration score of 4, and a risk score of 4.

Figure 6 is an example of a solution attempt that makes use of an assumption to answer the surface area problem. In this example, the student states that

"If the cuts were evenly made, the eighth block should be the same size as the 7th block." Using this assumption they they arrive at a solution of 28 units^2. This submission received an originality score of 3, a flexibility score of 3, a visualization score of 2, an elaboration score of 3, and a risk score of 3.

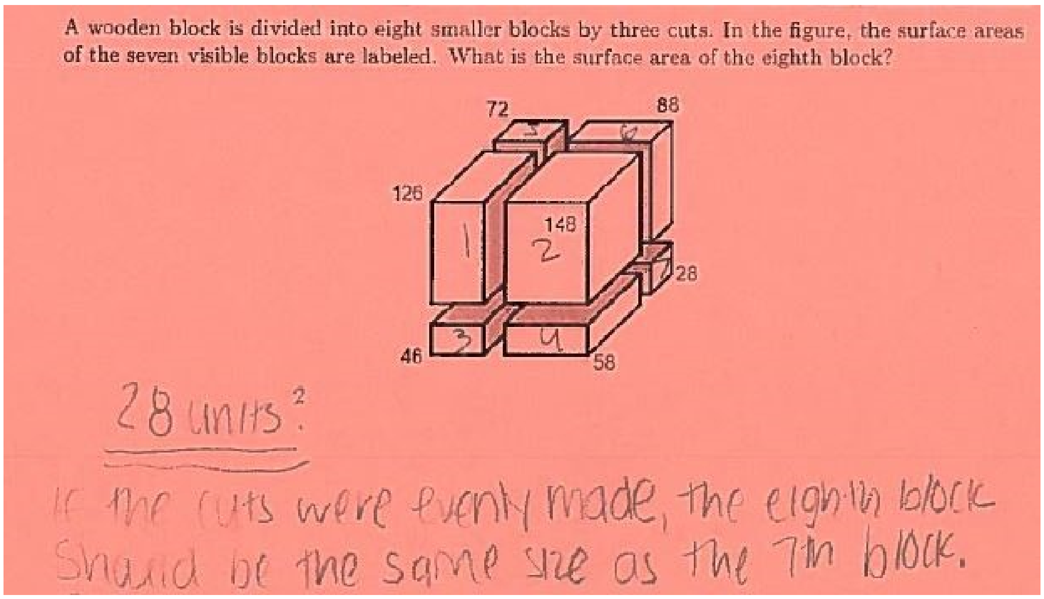


Figure 6: An attempt of the surface area problem where the student makes an assumption about the figure.

Figure 7 is an example of a solution attempt that makes use of proportions to answer the surface area problem. In this example, the student sets up several proportions involving the missing surface area and solves all of them. Seeing multiple values around 23, they they arrive at a solution of about 23. While the problem can not be solved exactly correctly, the proportion solution is one of the most efficient methods of arriving at a very good estimate that we saw. This submission received an originality score of 4, a flexibility score of 4, a visualization score of 1, an elaboration score of 3, and a risk score of 5.

Figure 8 is an example of a solution that recognizes a pattern in the differences of the given surface areas and then applies it to find the missing surface area. In this example, the student states that "The difference between$148 - 126 = 22$, difference between $58 - 46 = 12$. Therefore, since for back $88 - 72 = 16$ so $28 - x = 6$." Solving for x they arrive at $SA = 22$.

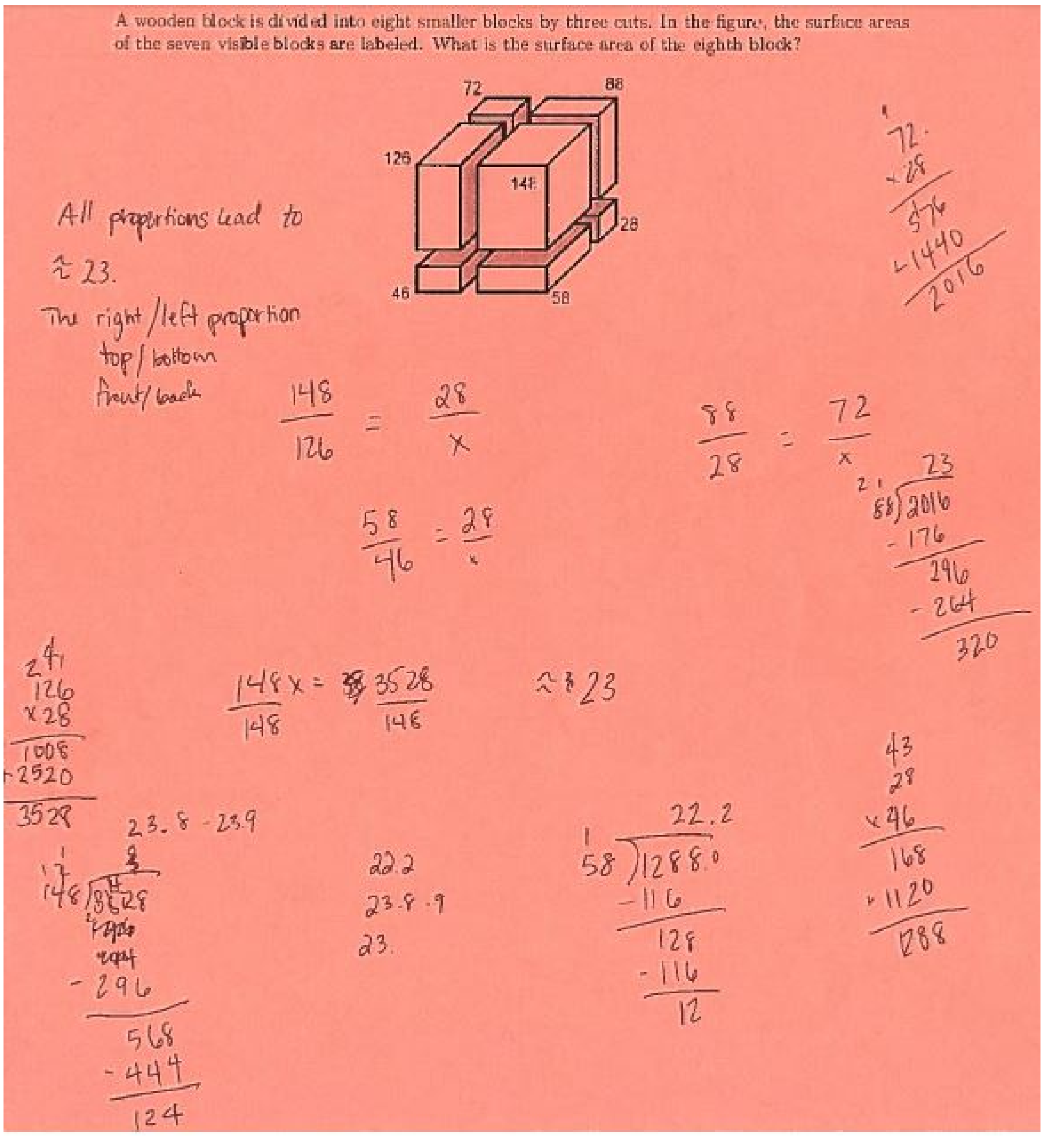


Figure 7: An attempt at the surface area problem using proportions.

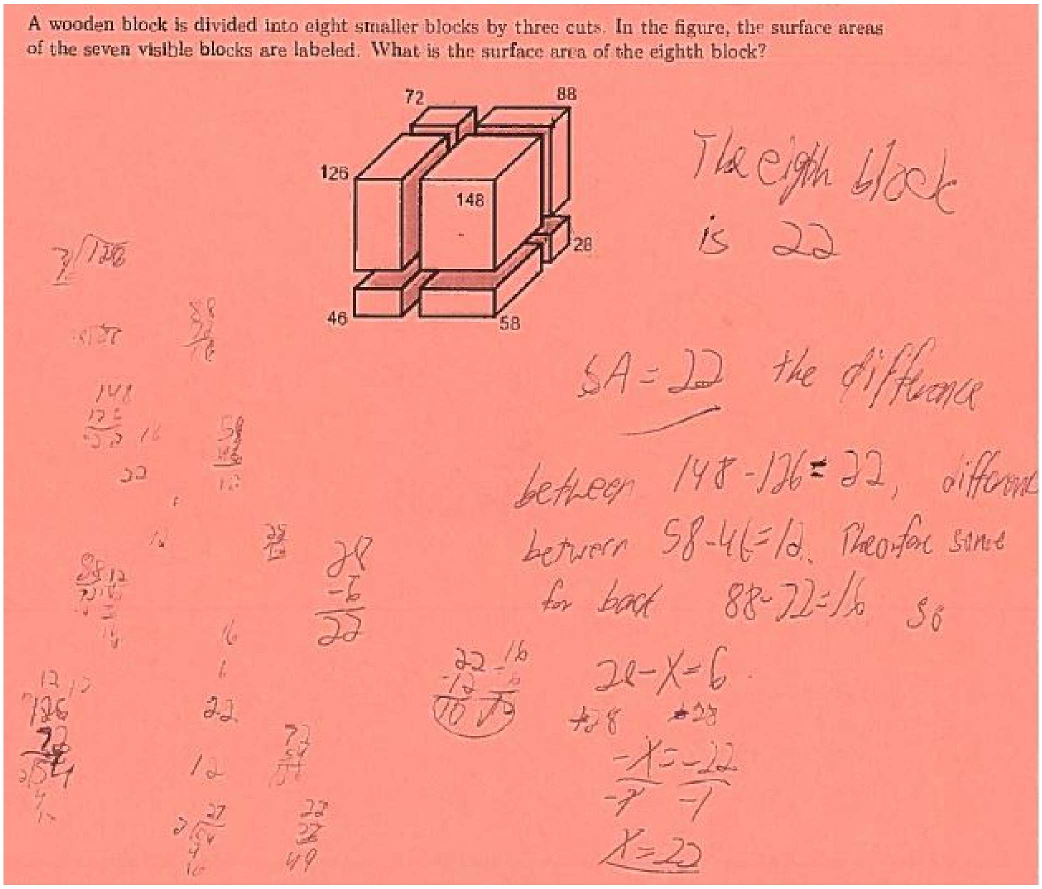


Figure 8: An attempt at the surface area problem using differences between related blocks.

While this student did not do the best job of explaining their process, they do manage to get their point across and their process is a sound one that reliably delivers the correct answer. This submission received an originality score of 5, a flexibility score of 5, a visualization score of 1, an elaboration score of 4, and a risk score of 5.

Figure 9 is an example of a submission that leaves the evaluator somewhat frustrated and wondering about the student's problem-solving technique. While the student arrives at the correct answer of 22 they record only a single thought about the missing surface area belonging to the smallest box. This submission received an originality score of 1, a flexibility score of 2, a visualization score of 3, an elaboration score of 2, and a risk score of 3.

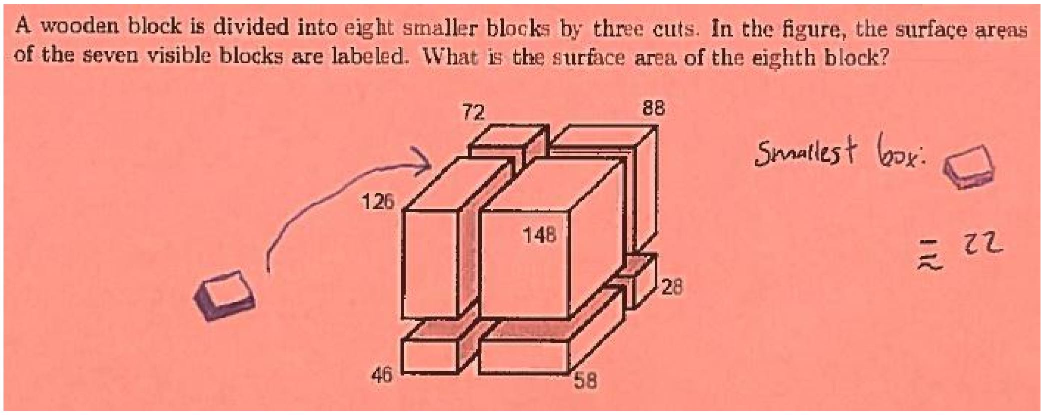


Figure 9: An attempt at the surface area problem with no communicated thought process.

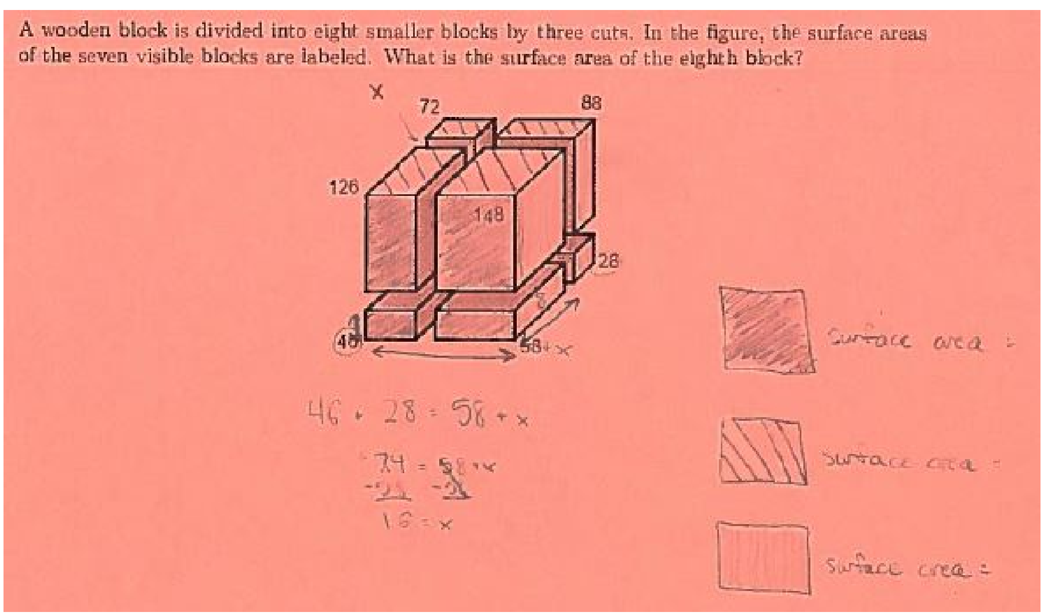


Figure 10: An attempt at the surface area problem that displays a lot of potential insight.

Figure 10 is an example of a student showing a lot of potential insight to the surface area problem without following through to a solution. In this example, the student sets the sum of the diagonal blocks on the base of the cube equal to one another. If the student had extended this reasoning to the entirety of the block they would have completed perhaps the most insightful solution seen. As it is, their observation is incomplete and without any elaboration providing insight to their thought process we have no way of knowing if they were truly doing anything more than adding a pair of given surface areas and setting them equal to the sum of a given surface area and the missing surface area randomly. They arrive at an answer of $16 = x$. This submission received an originality score of 2, a flexibility score of 4, a visualization score of 3, an elaboration score of 2, and a risk score of 2.

Figure 11 is an example of a submission that attempts to make use of an exhaustive brute-force method to solve the cheese problem. The student states "If the mouse goes around the center cube when eating adjacent cubes, I still don't think it will be possible to eat the last day. I figured this by looking at my model and trying to find out which way the mouse could go. I could not find a route that the mouse could take." While the student arrives at the correct answer that it is not possible, their method must be exhaustive and there is no evidence to believe that it is. This submission received an originality score of 1, a flexibility score of 3, a visualization score of 3, an elaboration score of 3, and a risk score of 5.

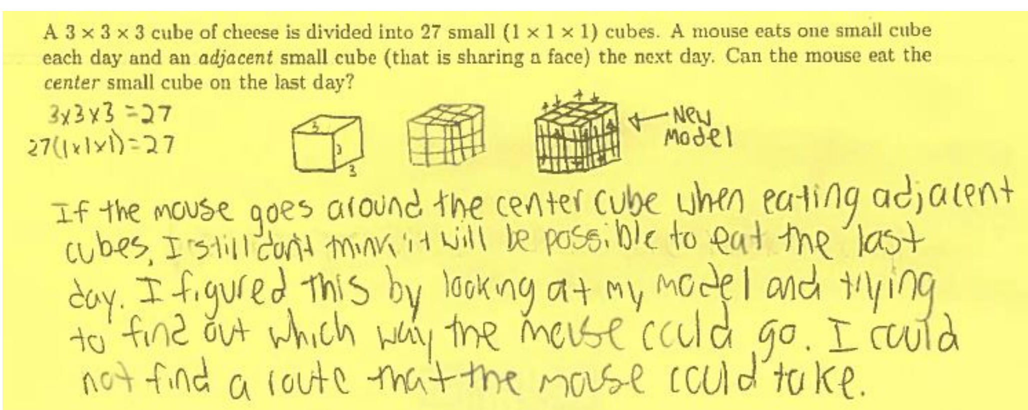


Figure 11: An attempt at the cheese problem using brute-force.

Figure 12 is an example of a submission that makes use of an assumption

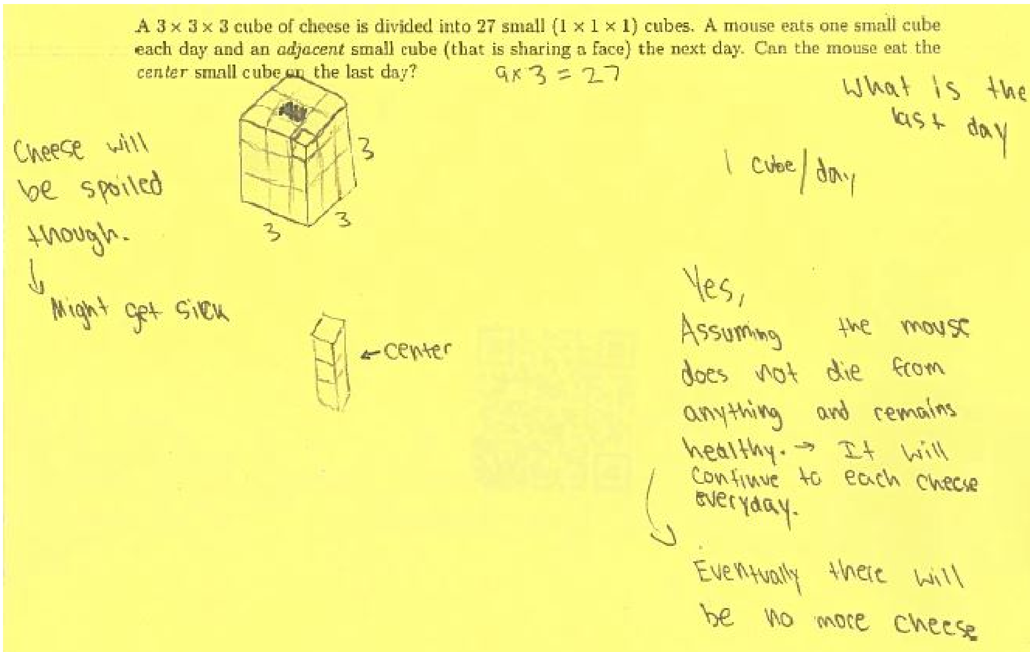
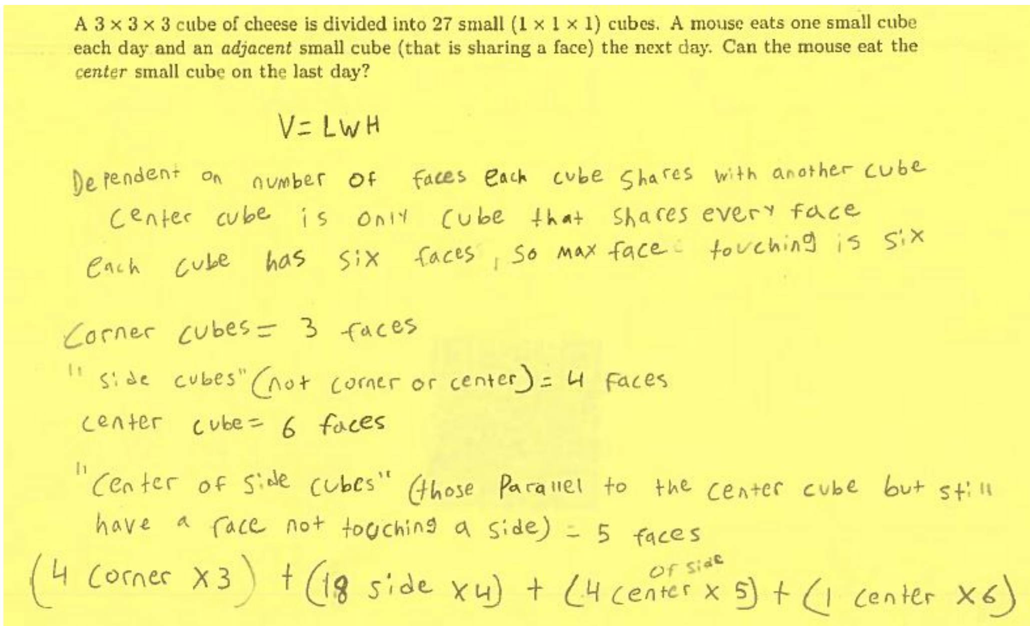


Figure 12: An attempt at the cheese problem using assumptions.

to answer the cheese problem. The student states that the "cheese will be spoiled though" which implies that the mouse "might get sick." They go on to answer "Yes, Assuming the mouse does not die from anything and remains healthy. → It will continue to eat cheese everyday. → Eventually there will be no more cheese." This is an instance of a time when the student seems to have misinterpreted or reinterpreted the question posed and answered their new question which is void of the depth and richness of the original problem instead. This submission received an originality score of 3, a flexibility score of 1, a visualization score of 3, an elaboration score of 3, and a risk score of 3.

Figure 13 is an example of a submission that attempts to use observations about the dimensions and volume of the cube to solve the cheese problem. After stating the volume formula for a rectangular prism, the student states "Dependent on number of faces each cube shares with another cube, center cube is only cube that shares every face, each cube has six faces, so max face touching is six." The student goes on to attempt to use those observations to arrive at an answer. Unfortunately, the student never answers the question



A 3 × 3 × 3 cube of cheese is divided into 27 small (1 × 1 × 1) cubes. A mouse eats one small cube each day and an *adjacent* small cube (that is sharing a face) the next day. Can the mouse eat the *center* small cube on the last day?

V= LWH

Dependent on number of faces each cube shares with another cube
center cube is only cube that shares every face
each cube has six faces, so max faces touching is six

Corner cubes = 3 faces
"side cubes" (not corner or center) = 4 faces
center cube = 6 faces
"Center of side cubes" (those parallel to the center cube but still have a face not touching a side) = 5 faces
(4 corner x3) + (18 side x4) + (4 center of side x 5) + (1 center x6)

Figure 13: An attempt at the cheese problem using observations about dimensions and volume.

posed. This submission received an originality score of 4, a flexibility score of 2, a visualization score of 1, an elaboration score of 4, and a risk score of 5.

Figure 14 is an example of a submission where the student considers the layers of the cube to solve the cheese problem. The student numbers the cubes – presumably determining an order in which the cubes should be eaten – in three ways. Then the student states "The mouse cannot eat the center small cube on the last day. each iteration leaves the center cube being eaten on the 2^{nd}" to last day." With this statement, the student not only answers the posed questions but goes a step further to communicate a bit of what seems to be a deeper understanding of the problem. This submission received an originality score of 4, a flexibility score of 4, a visualization score of 4, an elaboration score of 4, and a risk score of 5.

Figure 15 is an example similar to Figure 14; however, this student takes it a step further to explain how it could work after correctly answering the question as stated. The student states "The mouse cannot eat the last cube

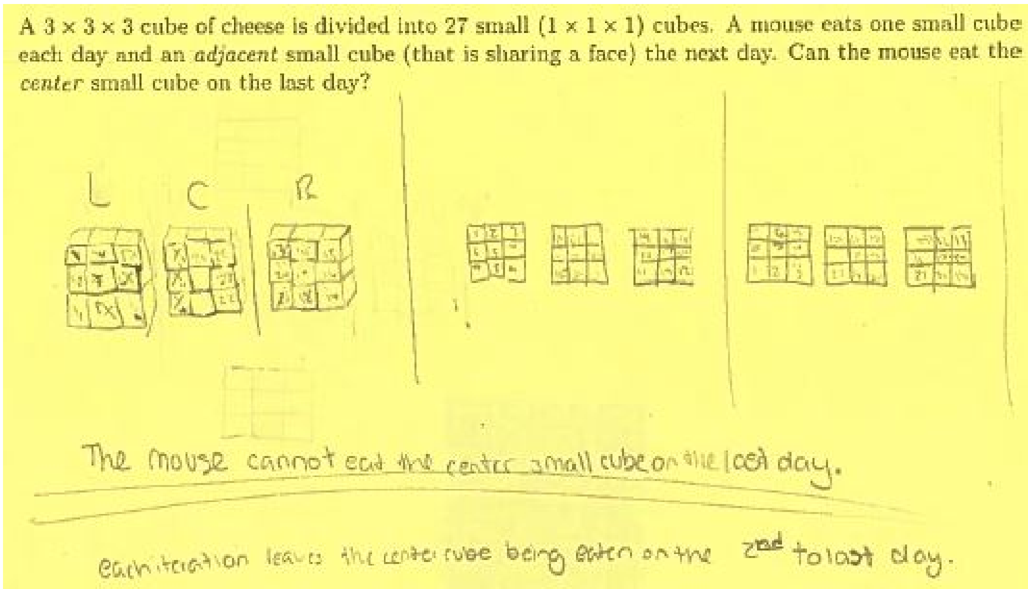


Figure 14: An attempt at the cheese problem by numbering cubes in layers.

on the last day because the last cube eaten would have to be diagonal to the center and therefore would not share a face. So, by the rule listed, it cannot be done. He would have to eat 3 a day for this to be possible, or the dimensions would have to be even numbers." This submission received an originality score of 4, a flexibility score of 3, a visualization score of 3, an elaboration score of 5, and a risk score of 5.

Figure 16 is an example of the most elegant solution we received the cheese problem. Another similar method that we also saw made the counting argument without the visual representation to accompany it. After shading a schematic of the layers in a checkerboard patterns that holds in three dimensions, the student makes a series of observations "14 shaded blocks; 13 non shaded; The mouse can't eat the same colored block 2 times in a row; center block will always fall into the group with 13 blocks to begin." The student then jumps to the conclusion that it is not possible. If they had made one more observation then they would have surpassed a correct answer and provided a rather elegant proof of their answer. This submission received an originality score of 5, a flexibility score of 3, a visualization score of 4, an elaboration score of 4, and a risk score of 5.

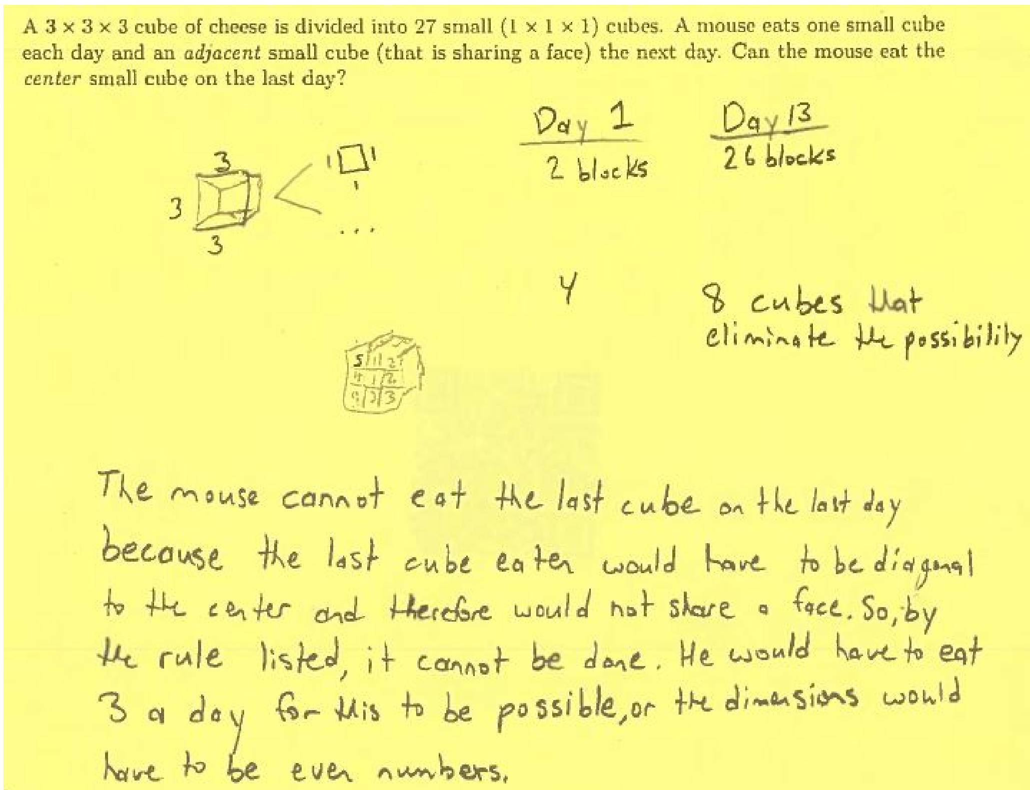


Figure 15: An attempt at the cheese problem that understands the next to last cube will be diagonal from the center.

Figure 17 is an example of a submission that leaves the evaluator wondering about the student's method to solve the cheese problem. The student draws a picture of a state that the cube of cheese cannot end up in given the constraints provided in the problem and then simply answers "No" without any explanation or work shown beyond the picture. While the student's problem solving may have been very creative, they provide the evlauator with no evidence of that creative problem solving. This submission received an originality score of 1, a flexibility score of 1, a visualization score of 2, an elaboration score of 1, and a risk score of 3.

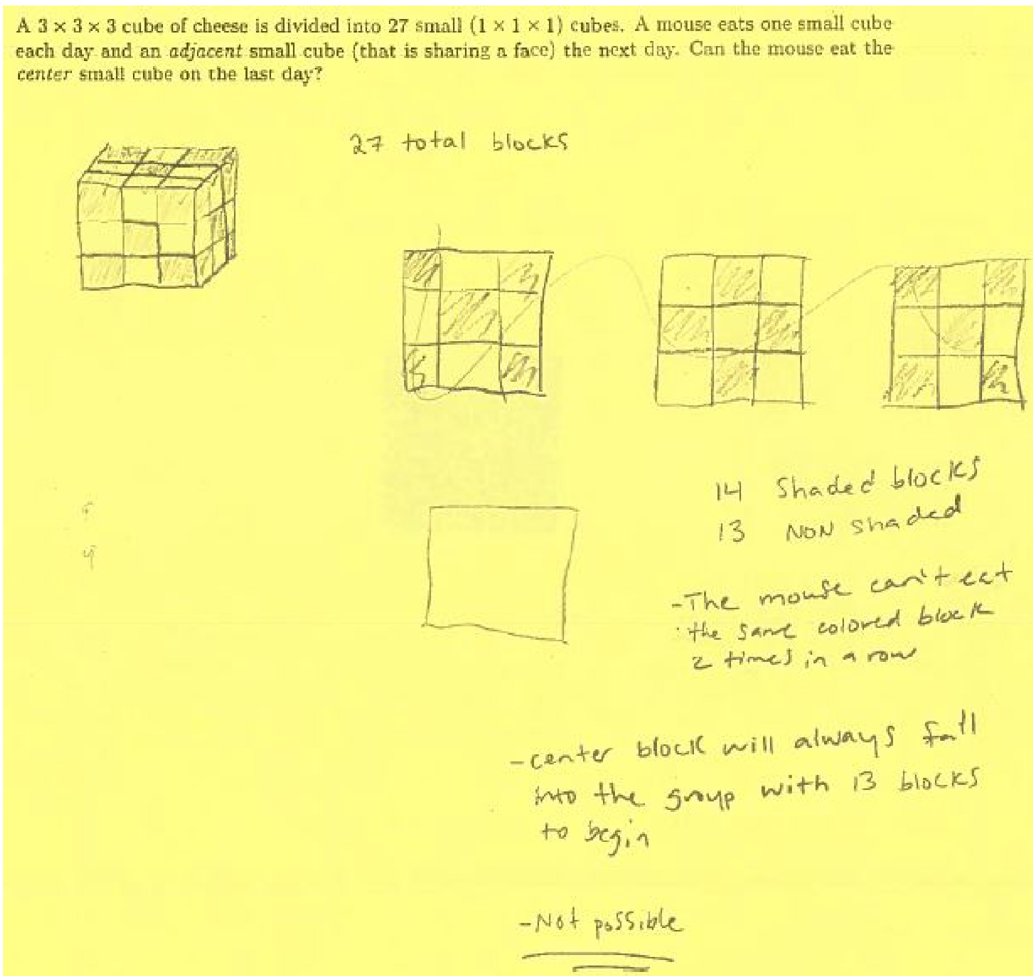


Figure 16: An attempt a the cheese problem using shading and a counting argument.

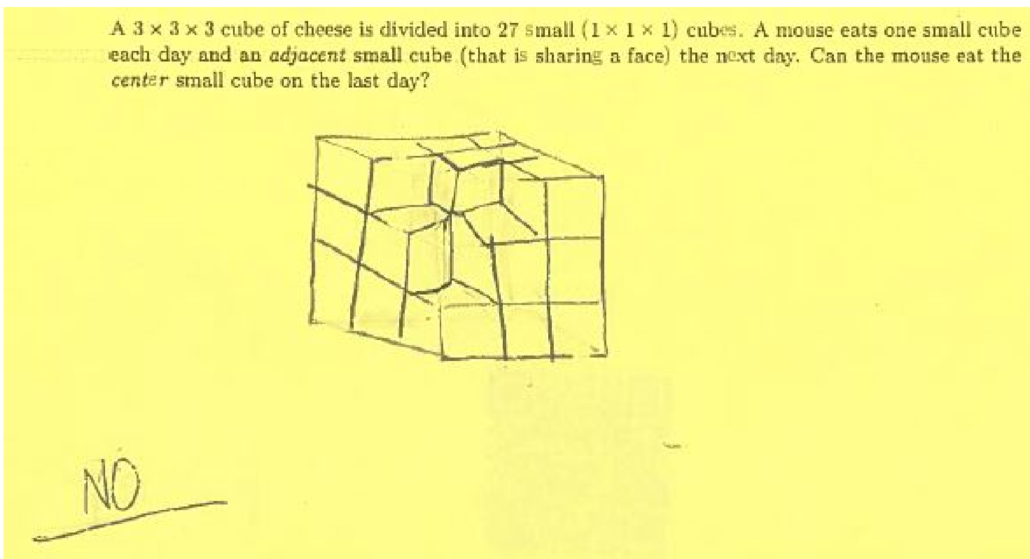


Figure 17: An attempt at the cheese problem with no communicated thought process.