Exploring Relationships Between Undergraduates' Plausible and Productive Reasoning and Their Success in Solving Mathematics Problems

Andrew Kercher[1] Kathryn Rhoads[1] James A. Mendoza Álvarez[1]
[1]The University of Texas at Arlington

This study examines how the use of plausible and productive reasoning in mathematical problem solving (MPS) influences student performance on non-traditional problems. Data comes from ten individual, task-based interviews with College Algebra students. In general, students who demonstrated high use of plausible and productive reasoning had a higher percentage of correct answers on interview tasks than their peers. We propose reasons why a student may use plausible and productive reasoning and still arrive at an incorrect answer; we also consider how a student may use suboptimal reasoning and reach a correct answer.

Keywords: mathematical problem solving, plausible mathematical reasoning, College Algebra

Schoenfeld (1985) indicated that possession of relevant mathematical knowledge, facts, algorithmic procedures, and other domain knowledge were not sufficient for student success in mathematical problem solving (MPS); students often fail at MPS for other reasons. The purpose of this study is to explore the relationship between entry-level undergraduates' MPS practices and the correctness of their answers to mathematical problems. In particular, we focus on undergraduate students enrolled in a College Algebra course to explore the following research questions: a) To what extent is the amount of plausible and productive reasoning a student exhibits related to their success in accurately solving mathematics problems? b) What factors may contribute to perceived discrepancies between the amount of plausible and productive reasoning a student exhibits and their success in accurately solving mathematics problems?

Theoretical Perspective

The research literature contains several definitions for a mathematics *problem* (e.g., Schoenfeld, 1992; Wilson, Fernandez, & Hadaway, 1993). In our work, we adopt Lester's (2013) definition that "… a problem is a task for which an individual does not know (immediately) how to get an answer …" (p. 247). We distinguish a problem from a mathematical *exercise*, which we consider to be a routine scenario for applying mathematical knowledge and skills (Schoenfeld, 1983). By our definition, a particular mathematical task may be a problem for some students and not others (Schoenfeld, 1985). The problems we discuss in this paper are aimed at the audience of entry-level university students.

The process of MPS has also been described and defined by several researchers, and Campbell (2014) analyzed 25 research articles focused on MPS to characterize the process. He categorized the explicit or implicit definitions of MPS in the reviewed articles. Álvarez, Rhoads, and Campbell (in press) revised and refined Campbell's initial categorization and identified five key domains of MPS.

- **Sense-making:** Identifying key ideas and concepts to understand the underlying nature of the problem. Attending to the meaning of the problem posed.
- **Representing/connecting:** Reformulating the problem by using a representation not already used in the problem or connecting the problem to seemingly disjoint prior knowledge. Using multiple representations or connecting several areas of mathematics (e.g. geometric and algebraic concepts).

- **Reviewing:** Self-monitoring or assessing progress as problem solving occurs, or assessing the problem solution (e.g., checking for reasonableness) once the problem-solving process has concluded.
- **Justifying:** Communicating reasons for the methods and techniques used to arrive at a solution. Justifying solution method(s) or approach(es).
- **Challenge:** The problem must be challenging enough from the perspective of the problem solver to engage them in deep thinking or processes toward a goal, "without an immediate means of reaching the goal" (Wilson et al., 1993, p. 57).

We also draw on Lithner's (2000) characterization of undergraduate students' reasoning as they solve mathematical tasks. Lither argued that students' reasoning could be *plausible* or based on past *experiences*. In using plausible reasoning, students rely on "the mathematical properties of the components involved in the reasoning" (Lithner, 2000, p. 167). Formal proof is an example of plausible reasoning, although Lithner's definition also allows for less-rigorous reasoning, as long as it relies on mathematical principles to reach a conclusion. By contrast, reasoning based on experiences relies on the student's past experiences in mathematics class or elsewhere. In this type of reasoning, students draw conclusions based on what they have observed or experienced in the past, without connection to the underlying mathematical principles. For example, when given a quadratic expression as part of a problem, students may assume it can be factored if they have worked primarily with factorable quadratics in the past. Lithner illustrated how plausible reasoning was sparser than experienced reasoning, but emphasized that reasoning from past experiences can be a useful strategy in MPS when students also use plausible reasoning in the process.

In this paper, we describe undergraduate students' MPS in terms of both the MPS domain they employ (Álvarez et al., in press) and the type of reasoning that underlies the use of that domain (Lithner, 2000). For example, a student may make a choice in representing a problem (representing/connecting domain) based on sound mathematical principles, or they may use a representation based on their past experiences.

Research Methodology

Setting

The data for this study comes from the Mathematical Problem Solving Item Development Project, in which we aim to develop efficiently-scored survey items assessing undergraduate students' MPS in each of the domains described by Álvarez et al. (in press): sense-making, representing/connecting, reviewing, justifying, and challenge. As part of the project, we use an MPS survey consisting of five mathematics *problems* and a number of associated *items*, with each item linked to one MPS domain. The problems were designed to be open-ended and appropriately challenging for undergraduates, but do not require knowledge beyond secondary-school algebra. A sample problem is shown in Figure 1. (For additional survey information, see Álvarez et al., in press.)

> Fun Golf, a local mini-golf course, charges $5 to play one round of mini-golf. At this price, Fun Golf sells 120 rounds per week on average. After studying the relevant information, the manager says for each $1 increase in price, five fewer rounds will be sold each week. To maximize revenues, how much should Fun Golf charge for one round?

Figure 1. Sample problem from MPS survey.

The MPS survey was administered during the fall 2016 semester at a large, urban university in the southwest United States. The survey was administered in College Algebra and Calculus courses designed for undergraduates intending to major in a STEM degree. A pre-test version of the survey was completed by 492 College Algebra students during class time at the beginning of the semester.

Participants

Participants for this study were 10 students chosen from the pool of 492 College Algebra students who completed the MPS pre-test in fall 2016. Interview invitations were sent to various students in an attempt to interview a diverse group of students in terms of gender and their performance on the pre-test. However, due to a limited number of responses to invitations, participants mostly represented a convenience sample. Of the students interviewed, four were male and six were female. All except one were 18 years old. All except one were STEM majors. All had completed a previous mathematics course at a level beyond second-year school algebra, graduated high school in spring of 2016, and were now enrolled in their first year of university studies. Eight had their last mathematics course within the last year. Pseudonyms linked to participant identification numbers were assigned to the students interviewed.

During the fall 2016 semester, each of the 10 participants took part in an individual, one-hour interview with one of the researchers. An interview consisted of completing three problems, during which the student was asked to explain their work while solving each problem[1]. Of the three problems, one was new to the interview participant. The other two problems were selected from the MPS pre-test the student had already completed. After solving one of these older problems, the participant was given a chance to review their original work and explain any differences in approach. All interviews were video-recorded and later transcribed, and all physical work was collected for analysis.

Data Analysis

To analyze the data, we used thematic analysis (e.g., Braun & Clarke, 2006; Nowell, Norris, White, & Moules, 2017). Interviews were conducted to understand the MPS practices of entry-level undergraduates, and we were particularly interested in the five MPS domains specified in the theoretical perspective. As such, a preliminary coding framework for the interviews was designed to identify only usage of the MPS domains (e.g., Miles & Huberman, 1994). While coding, we discovered that identifying only instances of MPS domain usage was insufficient for describing the subtle differences in student work. The coding scheme was then adjusted using both inductive and deductive approaches to incorporate an array of subcategories within each MPS domain (e.g., Nowell et al., 2017). In particular, each specific instance of an MPS domain was simultaneously assigned two sub-categorizations intended to describe its *utility* and *origin*, respectively.

The utility of an instance of MPS was further coded as *productive*, *conditionally productive*, or *non-productive.* Productive use of an MPS domain involved using that domain in a way that brought the student closer to an acceptable answer or that helped them avoid an unacceptable answer. Non-productive use of a domain corresponded to the negation of productive use. Conditionally productive MPS corresponded to work that led to a correct answer in the interview, but may lead to incorrect answers on other, similar questions.

---

[1] Students also responded to corresponding MPS assessment items under development. Items are not discussed in this paper, but are described in Álvarez et al. (in press).

Along the other axis, we further granulated instances of MPS by examining the origin of the student's MPS reasoning process. We adapted Lithner's (2000) classification of reasoning styles as either *plausible* or based on past *experiences*—with the same dichotomy applied to MPS domain usage. We also made use of a third category, *indeterminate*, for cases when the origin of a student's reasoning could not be determined.

For example, Amy was able to make a rough sketch of three parabolas that helped her to make progress toward solving one of the problems, and this was coded as productive representing/connecting using plausible reasoning. By contrast, when solving a different problem, Amy guessed that the graph of a relationship would look like the graph of either a cubic function or a linear function. This inference did not help her to work towards an answer, and it was not clear on what reasoning her conclusion was based. This excerpt was coded as non-productive representing/connecting with indeterminate reasoning. As a final example, when working on a problem involving revenue, Liz claimed that as the sales price increased, the revenue would increase to a point and then "it'll start going down because people will stop buying." In the problem, it was mathematically the case that the revenue reached a maximum and then decreased, but her conclusion is not generalizable to other situations. In addition, Liz's reasoning was not based on mathematics but rather her past experiences. Hence, this situation was coded as conditionally productive sense-making using experiential reasoning.

Transcriptions of the interviews were coded independently by at least two researchers. At multiple points in the coding process, the researchers compared excerpts of coding to refine the coding scheme and resolve conflicts. Once all coding was complete, the results were analyzed and collated, again resolving any remaining conflicts.

To gauge student success on the MPS survey problems, it was necessary to establish a grading scheme for assessing their work generated during the interview. Although each student addressed three unique problems, the third problem was often not attended to with as much detail or rigor as the first two problems. So, we elected to score only the first two problems by assigning each question 50% of the student's overall score for the interview; then, any problem that was comprised of more than one sub-problem divided its 50% equally among those sub-problems. For example, a student who completed the problems Fun Golf (a one-part problem) and Air Travel (a three-part problem) during the interview could earn 50% credit for correctly answering Ken's Garden and an additional 16.6% credit for each of the three parts of Air Travel they answered correctly.

We then considered the correlation between instances of MPS coded in the interviews and the score on the interview problems. We were also interested in the factors that may contribute to perceived discrepancies between the amount of plausible and productive reasoning a student exhibits and their ability to accurately solve these problems. To explore this, we revisited the coded data and searched for possible explanations for discrepancies, using an iterative approach to refine these explanations (e.g., Yin, 2009).

Results

Observable Correlations

We noted strong, positive linear correlations between students' interview scores and two separate, but related, metrics: instances of productive MPS based on plausible reasoning ($r = .813$; shown as Plaus/Prod MPS # in Table 1), and the percent of MPS instances that were both plausible and based on productive reasoning ($r = .807$; shown as Plaus/Prod MPS % in Table 1).

Table 1. Percentage of Plaus/Prod MPS corresponding to "interview score".

Participant	Score %	Total MPS #	Plaus/Prod MPS #	Plaus/Prod MPS %
Jill	0%	9	1	11%
Amy	0%	11	2	18%
Zoe	0%	8	2	25%
Dan	25%	11	2	18%
Sara	33%	20	5	25%
Liz	33%	13	5	38%
Ian	50%	7	1	14%
Matt	50%	14	5	36%
Bob	75%	14	6	43%
Kim	100%	15	9	60%

Although the strong linear correlations exist, we recognize that interview scores were more categorical than continuous and inconsistent among students. For example, Kim was given two problems that were each single prompts requesting one answer. Dan completed two problems that encompassed five total sub-questions. We also recognize that each student demonstrated a different number of discrete instances of MPS during their interview (Total MPS # in Table 1). For students who demonstrated a low overall frequency of MPS, the corresponding percent of plausible and productive MPS is also undesirably categorical. Ian and Zoe's interviews were examples of this flaw, and removing them results in a large increase in both r-values (to .940 and .942, respectively).

Taking these limitations into account, we were interested in possible explanations for why the percent of plausible and productive MPS used by a student may not have provided a direct prediction for the percent score they made on the interview questions. We discuss possible reasons in the following sections.

Plausible and productive reasoning in concurrence with incorrect answers

We now discuss possible reasons why a student may demonstrate a nonzero amount of plausible and productive MPS practices but still earn an especially low score on the interview problems. In general, it is sufficient to note that a problem often requires more than one instance of "good" MPS to arrive at a correct answer.

For example, Amy was solving a problem about two runners in a race. In her solving process, she revisited the problem statement and identified an error in her work, which is an example of productive reviewing using plausible reasoning:

> *Amy*: Alright, so looking at it… It just says that Brett finishes the 100 in 16 so that means that the 80 he did not complete in 16 so automatically I need to change that [erases mislabeled diagram]. So at this point if I don't understand it, I'll just take a guess.

However, as shown in the excerpt, although Amy was able to refer back to the problem text to identify and avoid a mistake, she was then unable to use good sense-making to correctly orient herself in a more productive direction, and ultimately she decided to "take a guess" at an answer.

Another example can be found in the interview with Zoe, who worked to solve the problem shown in Figure 1 regarding revenue at a mini golf course. After reading the problem, she made sense of the given conditions:

Zoe: Okay, so $5 for one round equals 120 rounds per week. And they're saying if they increase by $1, which could be $6 [per round], they will get 5 fewer rounds, which would be 115 rounds per week, and then they want to maximize their revenue, how much they bring in, so they have to charge a dollar decrease by $1 for $4 [per round] which would give them 125 rounds per week.

Zoe demonstrated productive sense-making using plausible reasoning, both when describing the effects of changing the price for a round of golf and when correctly attending to the meaning of the word revenue. However, she then went on to display non-productive sense-making by incorrectly interpreting how to maximize the revenue, arriving at an incorrect answer.

Limited plausible and productive reasoning in concurrence with correct answers

We now propose possible reasons why a student can achieve an interview score significantly higher than the percent of plausible and productive reasoning they exhibit. First, non-productive or conditionally productive MPS need not lead to incorrect answers; and second, students who have false starts are able to later correct themselves through a combination of appropriately plausible and productive reviewing and sense-making.

Kim's work exemplified the first point. Her use of representing/connecting illustrates how our classifications of domain use may contribute to a misleading characterization of an approach. Kim displayed three unique instances of representing/connecting in her work across two problems. Each instance was plausible, but two were non-productive. These non-productive instances were "trivial" in that they did not explicitly lead to either a correct or an incorrect answer. For example, while working a problem involving the area of a rectangular garden, Kim drew a simple diagram representing the garden, which did nothing more than extract the relevant dimensional information from the problem text. This qualifies as representing/connecting and uses plausible reasoning, yet is non-productive because the diagram itself does not play a meaningful role in Kim's approach to the problem. Had Kim used the diagram to robustly model the situation, it would have been productive. But by drawing the diagram, Kim lowered her percent of plausible and productive MPS but still answered the problem correctly.

Kim provided another example, this time of conditional productivity, while working the Fun Golf problem (see Figure 1). Kim claimed, "Yeah. Cause I thought it would just keep going up, but I realized maximize and minimum would mean quadratic." This is an example of experienced and conditionally-productive reviewing. Kim reasoned about the behavior of the revenue function using her experiential association of the word "maximum" with the vertex of quadratic functions. The revenue function in Fun Golf does happen to be quadratic, but certainly not every optimization problem involves second-degree polynomials. Thus, this particular instance of MPS is neither plausible nor explicitly productive by our definition. Still, it contributed to Kim's eventual success in the Fun Golf problem by helping her assess her progress, lowering percent plausible and productive MPS but contributing to a correct answer.

Finally, we consider a student who commits to an incorrect approach to a problem until recognizing a mistake and correcting herself with plausible and productive MPS. Sara worked on the Air Travel problem, as shown in Figure 2.

A commercial jet is flying from Boston to Los Angeles. The approximate distance in miles between Los Angeles and the jet can be found using the function $g(t) = -475t + 2650$, where t is the number of hours the jet has been flying. (i) Find a function, f, modeling the plane's distance from Los Angeles (in miles) in terms of *v*, where *v* is the number of minutes the plane has been flying. (ii) How far has the plane flown after 12 minutes?

Figure 2. Air Travel problem.

When beginning part (i) of the Air Travel problem, Sara remarked, “So since *v* is the number of minutes, and then this one, *t*, is the number of hours, we'd have to do *v* times 60.” This excerpt is an example of plausible representing/connecting, because Sara drew a connection between the units using the variables given in the problem text, but the MPS is non-productive, because the relationship she described is not correct. However, Sara soon made the following realization when using her non-productive MPS as a basis for her approach to part (ii):

Sara: f is equal to -475, 60 times 12 plus... [mumbling] ...is 720 minutes. Hmm. [using calculator] Mmkay, what I—sorry, I didn't write it down, what I was doing was trying to see-- I think you have to divide it by 60. Because you're dividing the minutes into the hours... And so, I just checked seeing what 12 divided by 60 was to see if it was a fifth and it is a fifth, so. It would-- this would be v over 60, I would think.

This excerpt exemplifies plausible and productive reviewing. Sara realized that 60 times 12 is 720 minutes, not 720 hours, as she had previously implied. She used this insight to evaluate an alternative—that minutes divided by 60 equals hours—and used a computation to judge that this relationship is more reasonable. In this way, Sara leveraged her initial non-productive MPS toward a correct answer by eliminating an incorrect possibility. A student who is often engaged in non-productive MPS may eventually arrive at a correct answer but with surprisingly low percent of plausible and productive MPS.

Discussion and Implications

Our results suggest that the amount of plausible and productive reasoning that undergraduate students use in solving mathematical problems may strongly correlate to their success on such problems. However, we also provide reasons why a student’s plausible and productive reasoning would not need to be extremely high to answer problems correctly and why a student may use plausible and productive reasoning yet answer problems incorrectly. As shown in our data, reasoning based on past experiences is not necessarily detrimental to the solving process, and in fact, as Lithner (2000) found, reasoning based on past experiences may be helpful in MPS. In addition, non-productive solving paths do not necessarily lead to incorrect answers.

Nonetheless, as Lithner (2000) cautioned, students often generalize from the examples and exercises they see in mathematics class, sometimes inappropriately. This can lead to over-application of experiential reasoning that is not balanced by plausible reasoning. Solving more non-routine problems may offer an opportunity for students to see that experiential reasoning is not always useful. Undergraduate mathematics instructors may want to ask students to explain their reasoning as they solve such problems and be attuned to the types of reasoning being used.

Future research could explore whether the trends that we observed hold for a larger sample. A larger sample could also illustrate whether certain MPS domains are more often backed by plausible and productive reasoning (or experienced or non-productive reasoning).

Acknowledgement

Partial support for this work was provided by the National Science Foundation (NSF) Improving Undergraduate STEM Education (IUSE) program under Award No. 1544545. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

Álvarez, J. A. M., Rhoads, K., & Campbell, R. C. (in press). Toward designing and developing Likert items to assess mathematical problem solving.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77-101.

Campbell, R. C. (2014). *Reconciling secondary inservice teachers' definitions of problem solving with those found in the research literature* (Unpublished master's thesis). The University of Texas at Arlington, Arlington, TX.

Lester, F. (2013). Thoughts about research on mathematical problem-solving instruction. *The Mathematics Enthusiast, 10*(1&2), 245-278.

Lithner, J. (2000). Mathematical reasoning in task solving. *Educational Studies in Mathematics, 41*(2), 165-190.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd edition). Thousand Oaks, CA: Sage Publications.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*, 1-13.

Schoenfeld, A. H. (1983). The wild, wild, wild, wild, wild world of problem solving: A review of sorts. *For the Learning of Mathematics* 3, 40–47.

Schoenfeld, A. H. (1985). *Mathematical Problem Solving*. Orlando, FL: Academic Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334-370). New York: MacMillan.

Wilson, J. W., Fernandez, M. L., & Hadaway, N. (1993). Mathematical problem solving. In P. S. Wilson (Ed.), *Research ideas for the classroom: High school mathematics* (pp. 57-78). New York, NY: MacMillan.

Yin, R. K. (2009). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.