

ENCOURAGING SIXTH-GRADE STUDENTS' PROBLEM-SOLVING PERFORMANCE BY TEACHING THROUGH PROBLEM SOLVING

Jonathan D. Bostic
Bowling Green State University
bosticj@bgsu.edu

Stephen J. Pape
John Hopkins University

Tim Jacobbe
University of Florida

Abstract

This teaching experiment provided students with continuous engagement in a problem-solving based instructional approach during one mathematics unit. Three sections of sixth-grade mathematics were sampled from a school in Florida, U.S.A. and one section was randomly assigned to experience teaching through problem solving. Students' problem-solving performance and performance on a unit test were analyzed. The intervention had a positive effect on students' problem-solving performance whereas the comparison group experienced no changes. ANCOVA analyses suggest that intervention students solved more problems on the posttest than their peers, but the comparison group outperformed the intervention group on the unit test.

Problem solving has long been a central theme within mathematics education, the importance of which is seen in mathematics standards around the world. To name a few, much of the United States adopted the Standards for Mathematical Practice and Standards for Mathematical Content (National Governors Association, Council of Chief State School Officers [NGA, CCSSO], 2010), Australia draws upon the Australian Mathematics Curriculum (Australian Curriculum, Assessment, and Reporting Authority, 2014), and Japan follows courses of study in mathematics (Ministry of

Education, Culture, Sports, Science and Technology, 2007). The National Council of Teachers of Mathematics (NCTM, 1980, 1989, 2000, 2006, 2009) has consistently advocated for problem solving as part of day-to-day mathematics instruction, with the rationale that solving problems is central to doing and learning mathematics (Ball, Ferrini-Mundy, Kilpatrick, Milgram, Schmid, & Schaar, 2005; Davis, 1992; Kilpatrick, Swafford, & Findell, 2001; Lester, 1994). The overarching goal of the present study is to describe an instantiation of one type of problem-solving instruction within a middle school mathematics classroom and explore students' outcomes compared to their peers who experienced traditional teacher-led explicit instruction. We not only explore this problem-solving instruction and its outcomes but also problematize a problem-solving approach to mathematics instruction.

Problems and Exercises

Problem solving involves a problematic task, which offers a goal for the problem solver to accomplish, but the means for achieving the goal are not readily apparent (Lesh & Zawojewski, 2007; Schoenfeld, 2011). A problem requires the problem solver to make sense of a problem situation and to make decision about a path to solution, which directs an individual toward the desired goal (Schoenfeld, 2011). Problem solving can be challenging because a solution is uncertain and/or unknown to the problem solver. Problems are distinct from exercises, which have their place in instruction. Exercises provide students a context in which they might develop efficiency with a known procedure thus improving their procedural competence (Kilpatrick et al., 2001). Mathematics teaching that heavily relies on exercises, however, does not support students' problem-solving outcomes (Kilpatrick et al., 2001; NCTM, 2009).

Framing Teaching Through Problem Solving

Mathematics instruction frequently separates problem solving from daily mathematics teaching (Hiebert et al., 1996). This practice of separating the two encourages the notion that learning mathematics and learning to solve mathematics problems are distinct from one another (Hiebert et al., 1997; Hiebert et al., 1996; Hiebert & Wearne, 2003; Lambdin, 2003). A major instructional concern is how to integrate problem solving within daily mathematics teaching. Three distinct approaches to problem-solving instruction, teaching *about*, *for*, and *through* problem solving, have been discussed in the research literature (Schroeder & Lester, 1989). Teaching about problem solving usually involves heuristic instruction. Teaching for problem solving focuses on teaching students mathematics procedures with the intention that they apply this knowledge to solve problems. Teaching

through problem solving (TTPS) involves teaching mathematics concepts through problem-solving contexts, provides opportunities for students to develop higher-level thinking during mathematical problem solving, and takes place in an inquiry-oriented learning environment (Hiebert & Wearne, 2003; Lambdin, 2003; Schroder & Lester, 1989). TTPS typically begins with a word problem that addresses one or more mathematics concepts and has the potential to engage students in complex forms of reasoning. These problems typically have characteristics of high cognitive demand (i.e., rich) tasks (see Stein & Smith, 1998). This approach differs from teaching about and for problem solving on conceptual and procedural levels. TTPS instruction encourages students to learn mathematics without stripping away contexts such as those found in realistic settings. Students must make sense of the problem's situation and the underlying mathematics concepts and procedures to solve these problems. While teachers may encounter pedagogical and mathematical challenges for TTPS, problems encourage the greater goal that mathematics is a way to help students make sense of their world rather than a set of procedures to master (Verschaffel, Van Dooren, Greer, & Mukhopadhyah, 2010). Students engaged in TTPS have opportunities to develop problem-solving abilities and fluency (Sigurdson, Olson, & Mason, 1994). It is hypothesized that this growth in problem solving and fluency assists learners in building connections between concepts and procedures and developing greater adaptive reasoning for effectively and efficiently executing procedures at appropriate moments (Sigurdson et al., 1994). This is a hypothesis because such a claim stems from classroom-based research, which is inherently complex with a multitude of factors.

It is not possible to easily separate an intervention's effects attributed to the instructor, instructional method, tasks, and learning environment on students' outcomes when engaging in classroom-based research (Ridlon, 2009; Sigurdson et al., 1994; Verschaffel & De Corte, 1997; Verschaffel et al., 1999; Verschaffel et al., 2010). Hence the intervention in this study, TTPS, is defined as an integration of these components. These components are described more fully later. The present study aimed to extend the prior research with three objectives. First, we describe TTPS through vignettes from a month-long unit. Second, we investigated sixth-grade students' problem-solving performance and performance on a unit test following an instructional intervention that utilized TTPS. Problem-solving performance was characterized as the number of correct responses to developmentally appropriate word problems. The unit test measured students' knowledge about focal topics during the unit of instruction (i.e., rates, ratios, and data analysis). Third, we compared intervention students' outcomes with their peers who experienced their typical teacher-led explicit mathematics instruction.

Problem-Solving Instruction

Prior Research on Problem-Solving Instruction

Several studies across the world have explored students' problem-solving performance and provide support for the present investigation. Nearly thirty years ago Charles and Lester (1984) explored the impact of supplementing everyday mathematics instruction with 10-25 minutes of problem-solving experiences in U.S. classrooms. Fifth- and seventh-grade students who experienced this supplemental instruction had more positive problem-solving outcomes when compared to their peers who experienced traditional teacher-led explicit instruction focused on procedures. Sigurdson and colleagues (1994) compared students' outcomes after experiencing three types of instruction in Canadian classrooms: (a) traditional procedure-focused explicit instruction, (b) an inquiry approach that involved a focus on connections between mathematical concepts and procedures, and (c) an inquiry approach supplemented with 10 minutes of daily problem-solving work. Students in the latter groups significantly outperformed those in the first group on a test measuring general mathematics content knowledge and had significantly better problem-solving performance. These studies support the conclusion that mathematics instruction *supplemented* with problem-solving instruction focusing less on procedures leads to improved problem-solving outcomes and positive growth in mathematics content knowledge. The authors also raise the important question regarding the impact of *integrating* problem solving within mathematics teaching, rather than simply including problem solving as a supplement to this instruction (Sigurdson et al., 1994).

Verschaffel and De Corte (1997) conducted a teaching experiment in Flanders with 10-11 year olds that responded to this question. Problem solving and mathematics content instruction were integrated instead of supplementing mathematics content instruction with problem solving. Their goal was to explore whether students might give more realistic (not necessarily correct) solutions to problems after learning about a problem-solving model and solving realistic problems in a supportive learning environment meant to foster student-to-student discourse. Daily instruction lasted two-and-a-half hours over five lessons. Participants in the teaching experiment provided more realistic responses on the problem-solving tests than their peers in a traditional learning environment. While this intervention was modest in its duration, students' problem-solving performance improved following a short period of time engaged in mathematics instruction that integrated problem solving and content (Verschaffel & De Corte, 1997). Finally, in a study most similar to the present study, Verschaffel and colleagues (1999) developed, piloted, and implemented an instructional program for Dutch fifth-grade students to examine a program aimed at helping

learners employ a metacognitive strategy for solving mathematics word problems. Four sections of fifth-grade mathematics classes experienced 20 problem-solving lessons over a four-month period while a group of seven comparison sections experienced typical teacher-led mathematics instruction. Instruction was guided by three “pillars” of a successful mathematics learning environment (Verschaffel et al., 1999, p. 202): “(a) realistic, complex, and open problems, (b) independent as well as small- and whole-group instruction, and (c) supportive classroom expectations for engaging in mathematics.” Similar to the present study, the researchers administered a pre- and posttest that had similar problems across both measures, as well as an achievement test measuring general mathematical knowledge and skill. Students in both groups improved their problem-solving performance, but the intervention group made greater gains on the problem-solving measure and outperformed the comparison group on the achievement test. Based on these studies of problem-solving instruction, the present investigation sought to examine outcomes for students engaged in TTPS instruction for approximately 20 lessons implemented consecutively.

The Current Study

This exploratory mixed-methods investigation examined the effects of TTPS on sixth-grade students’ performance on a problem-solving measure and unit test. An embedded design mixed-methods approach was selected for this study because of the study’s aim and nature of the research questions. The embedded mixed-method design allows researchers to unpack statistical findings with qualitative evidence and concurrently, qualitative evidence is supported by quantitative results (Cresswell, 2012).

The first research question focused on within-group differences whereas the second and third questions addressed potential between-group differences. Three research questions guided the present study. (a) What is the impact of the intervention on students’ performance on a test of word problems? (b) Does performance on a test of word problems differ between students from the intervention and comparison groups? (c) Does performance on a teacher-constructed unit test following TTPS instruction differ between students from the intervention and comparison group? In addition, one objective of this study was to offer a description of TTPS as instantiated within this study. This description contextualizes the findings, gives voice to student-to-teacher and student-to-student interactions, and supports critically examining possible social aspects implicating students’ outcomes. This investigation values both the social factors (e.g., classroom environments, mathematical discourse, and interactions between students as well as their teachers) and cognitive factors (e.g., problem-solving ability and content knowledge) of education. Both factor types have been shown to impact students’ outcomes (Ridlon, 2009; Sigurdson et al., 1994; Verschaffel et al.,

1999). Therefore the findings from the present study are meant to spur further conversations about teaching and learning mathematics, teaching and learning problem solving, and fostering students’ mathematical proficiency.

Method

Setting and Participants

Students came from three sixth-grade mathematics sections taught by the same teacher within a K-12 school that represented the diversity of the state of Florida. Classes were arranged to meet on a modified block schedule, so classes met three times per week. All sections met for 60 minutes on Monday and 90 minutes on two additional days. One section was randomly assigned to receive the intervention. Eighteen students from the intervention classroom and 20 students from each comparison classroom volunteered for the study. These three sections met on the same days, one right after the other. None of the participants received services for a disability or were English Language Learners. Demographic information for the intervention and comparison group is provided in Table 1. More than half of the students identified themselves as white and approximately 20% qualified to receive free-or-reduced lunch (FRL).

Table 1. *Demographic information for participation.*

	Intervention Group ^a Number (Percent)	Comparison Group ^b Number (Percent)
Ethnicity		
White	11 (.61)	20 (.50)
Hispanic	3 (.17)	11 (.28)
African-American	3 (.17)	6 (.15)
Multiracial	1 (.03)	2 (.05)
Asian-American	0 (0)	1 (.02)
Gender		
Male	7 (.39)	18 (.45)
Female	11 (.61)	22 (.55)
Free-or-Reduced Lunch		
Yes	5 (.28)	7 (.18)
No	13 (.72)	33 (.82)

^a N = 18; ^b N = 40

Group Comparisons

We examined the intervention and comparison groups for comparability. To control for differences in prior mathematics instruction, students from the same teacher were assigned to the intervention and comparison conditions. Students’ gender, ethnicity, fifth-grade mathematics and read-

ing standardized test scores (i.e., Florida Comprehensive Assessment Tests (FCAT)), and FRL status were collected from students’ records by school faculty. There were no significant difference between the group’s standardized test scores on the reading FCAT, $F(1, 51) = .62, p = .44$, and mathematics FCAT, $F(1, 51) = .17, p = .68$ (see Table 2).

Table 2. *Group means and standard deviations related to fifth-grade FCAT scores*

	Intervention Group ^a	Comparison Group ^b
	Mean (SD)	Mean (SD)
Reading Scale Score ^a	330 (36)	340 (45)
Mathematics Scale Score ^a	350 (32)	354 (33)

^a N = 16; ^b N = 37

Chi-square analyses were also conducted to determine whether there were differences between the groups in terms of gender, ethnicity, and FRL status. No significant pre-intervention group differences existed, suggesting the groups had similar demographic characteristics.

Measures

Students completed three measures including a Problem-Solving Pretest, Problem-Solving Posttest, and a test measuring students’ knowledge related to unit-specific topics.

Problem-solving tests. Several steps were taken to create pretest and posttest problem-solving tasks. Initially, problems were translated from Verschaffel et al.’s (1999) problem-solving measures, adapted to suit students’ interests and prior knowledge, and revised to conform to American English grammar rules. Each problem-solving measure included five problems (see Appendix A for posttest items) that were matched for content between the pretest and posttest. Five items on each instrument exceeded the minimum number of items necessary to sufficiently measure a single construct (i.e., problem-solving ability; Ary, Cheser-Jacobs, Sorenson, & Razavieh, 2009). These tasks drew upon a variety of developmentally appropriate mathematics concepts and procedures that students should have experienced prior to the study as determined by several middle and elementary school teachers and mathematics educators. The teachers and mathematics educators also believed the situations embedded within the problems drew on realistic contexts. Finally, the group agreed that the problems could be solved using more than one approach.

A pilot study was conducted to determine how these measures function with sixth-grade students in the U.S. including the measure’s dimensionality, item parameters, and measure reliability (i.e., internal consistency and alternate-forms reliability). One hundred sixty-nine sixth-grade students from a nearby school district that had similar demographics to the present study’s setting participated in the pilot study (see Bostic, Pape, & Jacobbe, 2011). This

sample size for five items was adequate for 95% confidence in results with stable item calibrations in the 0.5 logit range (Linacre, 1994). Data were calculated using WINSTEPS Version 3.62.1 (Linacre, 2006). Rasch model analysis was employed for two reasons: to determine overall fit of the data to the Rasch model and then explore the relative item difficulty. Related to the first reason, item information for the pre- and post tests was explored, specifically looking at the mean-square (MNSQ) fit statistic of infit and outfit data. MNSQ assesses an item (as in this case) or person’s contribution to measurement productivity (Drouin, Horner, & Sondergeld, 2012). Perfect MNSQ is one unit, which is rarely found, and values far greater than two or less than 0.5 may potentially distort the measurement system (Linacre, 2002). MNSQ values are found in Tables 3 and 4.

Table 3. *Item Information for pretest*

Item #	Infit		Outfit	
	Mean Square	ZSTD	Mean Square	ZSTD
1	0.99	0	0.9	-0.2
2	0.99	-0.1	0.94	-0.4
3	1.02	0.2	1.14	0.6
4	0.78	-1.7	0.67	-1.4
5	0.94	-0.4	0.94	-0.2

Table 4. *Item information for posttest*

Item #	Infit		Outfit	
	Mean Square	ZSTD	Mean Square	ZSTD
1	1.26	2	2.03	2.2
2	0.93	-0.7	0.83	-1.1
3	1.01	0.2	0.99	0
4	1.04	0.4	1.04	0.3
5	0.8	-2.3	0.67	-2.2

Two items on the measures were slightly higher than two MNSQ units but were retained since they were reasonably close to the recommended threshold.

Rasch reliability is similar to traditional reliability and was computed for individuals. Reliability for the pretest and posttest was high, $\alpha = 0.96$ and $\alpha = 0.97$, respectively. This met the excellent threshold (Duncan, Bode, Lai, & Perera, 2003). Alternate-forms reliability was calculated using a correlation statistic. Results indicated that it exceeded the minimum to link scores across tests, $r = 0.97$ (Ary, Cheser-Jacobs, Sorenson, & Razavieh, 2009).

Rasch separation was examined to investigate how many distinct groups can be made based on respondents’ data. Rasch separation near two units suggests that only two groups of respondents can be formed: those who were

successful and those who were unsuccessful. Values greater than 3.00 are considered excellent (Duncan et al., 2003). The Rasch separation for the measures was exceptionally high, 4.81. Thus, respondents could be sorted into approximately four distinct groups. This psychometric-based evidence indicated that the measure adequately captured the construct, problem-solving ability, and did so reliably. Next, the authors explored item difficulty again using WINSTEPS Version 3.62.1 (Linacre, 2006).

Item difficulties characterize the likelihood a respondent will respond correctly to the item. Item difficulties are measured in logits. The item difficulty scale extends in both positive and negative directions but usually ends near three logits, with the average value set at zero logits. An item with a difficulty parameter of zero logits suggests a respondent has an equally likely chance to answer the item correctly or incorrectly. Items with values greater than one logit are considered moderately difficult for the average-ability respondent, whereas negative logit values are easier for the average-ability respondent. Problem-solving items are more cognitively taxing than rote mathematical exercises. Therefore it should be expected that item difficulties ought to be greater than zero logits. Results from investigating item difficulty suggested that items ranged from easier to moderate difficulty (see Table 5).

Table 5. *Item difficulties for problem-solving measures*

Item #	Item difficulty	
	Measure	Model Std. Error
Pretest		
1	-0.92	0.12
2	0.34	0.11
3	0.82	0.12
4	0.85	0.12
5	0.74	0.12
Posttest		
1	-0.96	0.12
2	0.29	0.1
3	0.37	0.1
4	0.55	0.11
5	0.39	0.1

Synthesizing these findings with the earlier psychometric evidence leads to the conclusion that psychometrically speaking; the five-item measures were working sufficiently.

Unit test. Students’ content knowledge related to rates, ratios, and data analysis was measured with a two-page unit test. Results from the unit test filled a needed gap in the literature. Previous studies explored general achievement or content outside the scope of the instruction; none have investigated students’ performance on a test covering only the content

addressed during the instructional intervention period. The classroom teacher adapted it from the assessment materials that accompanied the class textbook: *Big Ideas: Math 6* (Larson & Boswell, 2010). The test consisted of 16 short-answer items, of which five tasks required two or more correct responses to receive full credit. Twelve items asked students to rewrite ratios in simplest form and give the unit rate for a situation described in a verbal exercise. For example, one item asked students to write the statement “1200 calories in 3 liters” as a unit rate. The other tasks were focused on finding the mean, median, mode, and/or range of a data set. A sample data analysis task read “Find the median and mode(s) of the data set 4,6,5,4,4,5,4,8.” Students could earn up to 25 points on the unit test. Internal consistency was found to be acceptable, $\rho = .82$.

Procedure

All sixth-grade students received mathematics instruction in the same classroom resulting in an identical classroom layout and equitable access to materials (e.g., textbooks and manipulatives) across sections. During the teaching experiment, the first author became the instructor in one classroom while the classroom teacher continued her instruction in the two comparison classrooms. The classroom teacher was not present in the intervention classroom during the study.

Data collection. The pretest and posttest were administered during students’ regular mathematics class approximately one month apart. The instructor read the directions aloud to students prior to beginning the measure. Most students needed 30 minutes for the problem-solving measures. The classroom teacher administered the unit test, which took approximately 60 minutes to complete, in all three sections. A mathematics educator not affiliated with the study observed instruction, videotaped lessons, and took field notes in both classrooms on three randomly selected occasions during the second, third, and fourth week of instruction. These data were used to develop a description of the instruction in each of the classrooms.

Instruction: Standards, tasks, and questions. The following four sixth-grade benchmarks from the Next Generation Sunshine State Standards (NGSSS; Florida Department of Education, 2007) selected by the classroom teacher were the focus of instruction during the present study:

- MA.6.A.2.1 Use reasoning about multiplication and division to solve ratio and rate problems
- MA.6.A.2.2 Interpret and compare ratios and rates
- MA.6.S.6.1 Determine the measures of central tendency (mean, median, and mode) and variability (range) for a given set of data

- MA.6.S.6.2 Select and analyze the measures of central tendency or variability to represent, describe, analyze and/or summarize a data set for the purposes of answering questions appropriately.

Lessons in the intervention classroom conducted during block scheduled periods tended to follow this order: (a) check homework, (b) discuss issues related to homework, (c) complete introductory task, (da) individual work on one problem, (db) examine the problem with a partner or in a small group, (dc) discuss the problem with the entire class, and (e) and complete a concluding activity meant to stimulate reflection. Rich tasks help students see that mathematics is connected (Stein & Smith, 1998). In this manuscript, we use the terms problem and rich task synonymously because the problems aimed to promote connectedness within mathematics and were sufficiently complex to address features of high cognitive demand, as expressed in Stein and Smith's (1998) cognitive demand framework.

A brief outline of the process used to adapt problems from textbook tasks is provided here, but a more detailed description of the process may be found in Bostic (2012/2013). Creating word problems for each lesson began by examining the state-level standards and considering the mathematical relationships between them. Next, the instructor reflected on ways to turn tasks from resources such as their textbook (Larson & Boswell, 2010) into open-ended and complex word problems that drew on realistic contexts. Students were consulted about their interests and experiences to learn about contexts they perceived as realistic. The instructor explored the textbook and other classroom resources for tasks. Typically information was added to the textbook tasks to make it realistic. Finally, additional questions that required higher-level reasoning skills, such as analysis and evaluation, were included. A sample problem is shown in Appendix B.

Data Analysis

Students' performance on the pretest and posttest was scored as correct or incorrect and a sum was calculated. The researcher and a second coder randomly selected 20% of the tests and scored them independently. Interrater agreement was 100% ($r_{wg} = 1$; James, Demaree, & Wolf, 1984). The two coders scored the remaining 80% of the tests after reaching satisfactory interrater agreement. The classroom teacher scored the unit tests for the three sections. Each response was scored as correct or incorrect and was equally weighted.

A repeated measures t-test was used to determine whether the intervention improved students' problem-solving performance. The second research question was examined using ANCOVA to investigate differences between groups' problem-solving performance using the pretest as a covariate. Since the groups were similar in terms of their demographic charac-

teristics and the sample size resulted in limited power to test effects, demographic data were not included as covariates. The third research question examined differences between the intervention and comparison students' content knowledge as measured by the unit test. Data were analyzed using ANCOVA with the students' fifth-grade standardized mathematics score as a covariate.

Assumptions related to normality, homoscedasticity, linearity, and multicollinearity were also investigated. Results from examining the residual plots and statistical analyses provided sufficient evidence to justify use of ANCOVA. When there was a significant difference between groups, partial η^2 was calculated to examine the size of the effect.

Results

A Description of TTPS

A description of typical TTPS instruction in this study based on an examination of the videotapes and field notes is provided to frame the instructional intervention and student outcomes. Excerpts of classroom dialogue are provided to contextualize instructional aspects. Vignettes from three instructional days that exemplify unique aspects of the TTPS instruction implemented in this study are shared to frame the intervention.

Classroom environment. Two posters that reflected the expectations for classroom processes and questions students were to ask one another while solving problems were displayed in the classroom (see Appendix C). An agenda that indicated tasks to accomplish at the beginning of class as well as objectives for that day was projected. Students usually checked homework first and then began an introductory task.

Checking homework. Students examined their homework on their own and were asked to consider tasks, procedures, or concepts for discussion. After five minutes, students were asked to indicate homework tasks they wanted to review, and the instructor invited students to discuss these questions (e.g., "Does someone have ideas about this problem?"). One student was selected to explain his or her approach to solving the problem. Students frequently described aspects of the problem that were critical features for solving it. After this explanation, the instructor asked the student who originally posed the question whether it was resolved (e.g., "Does that make sense?" "Would you like him/her to describe it in another way?"). The instructor followed up by probing students for other ways to solve the same problem—typically one student shared an alternate representation or process. This continued until students' questions about the homework were resolved, which typically took 10 to 15 minutes.

Introductory task. Following the homework discussion, the instructor

reminded students to complete an introductory task, which was projected on the front whiteboard. Students usually completed it in 10 minutes or less. The instructor interacted with students one-on-one during this instructional part and asked them to share their thinking. Peers were selected to share their thinking based on Smith and Stein's (2011) suggestions for fostering productive classroom mathematics discussions: (a) complexity of students' ideas, (b) whether their solution strategies were concrete or abstract, and (c) the correctness of their answer.

Problem. Following the introductory task, the instructor posed a question and students' preferences that were related to the context of the problem the students would examine during class (e.g., "What was the name of the last restaurant where you ate pizza?"). Multiple students mentioned several local restaurants that they would later see within the context of the problem, and some shared their pizza preferences. Many shared that the cheapest pizza was not necessarily the best value. The instructor elaborated that the problem they would solve involved investigating pizza prices from various local establishments and distributed individual copies of the problem (see Appendix B).

The students were initially encouraged to work independently for a few minutes. Students were reminded that they could collaborate on the problem after working independently. Students usually spent 5 to 10 minutes on their own before forming small groups. When the instructor announced that independent work time was over, students formed pairs or triads on their own. After discussing the problem's context and goal, students discussed how to solve it. Small-group work typically began with peer-to-peer questions, such as "What do we need to do?" and "What do you think about this [problem]?"

For example, one student in a group of three started the conversation about the pizza problem with a question and then a second student proceeded to read the problem and share a goal.

S1: What's the goal of the task?

S2: [Reads task aloud.] What is the best value for a pizza?

S1: We have to find out how many slices there are. [Points to data in table.]

S3: It says costs of...

S2: [Pause while S1 and S3 reread problem. S2 works independently.] I found the lowest price!

At times, students challenged each other to justify their ideas (e.g., "Why are you doing that?"). With regard to this problem, groups of students shared that the best value for one pizza may not necessarily be the least expensive pizza. After agreeing on a mathematical representation, they carried out a set of procedures and interpreted the result. Students continued to share ideas in small groups for 15-25 minutes depending on the problem's complexity.

During small-group work, the instructor walked around the classroom observing students' work and responding to requests for assistance with questions such as "What do you think you're supposed to do?" and "What do you think is important in the problem?" The class reconvened to discuss the problem when most students were finished.

The instructor began the whole-class instruction by posing an open-ended question such as "What is going on in this problem?" or "What do we need to find?" Presenters typically discussed their mathematical representation, procedures used to solve the problem, thoughts about their problem solving, and answer(s). Some transcribed their work onto the whiteboard located at the front of the room whereas others used the document camera to project their work to frame their discussion. During another class session, students investigated a problem aimed to answer the question "What type of music is preferred by students in the sixth grade?" Students and the instructor discussed how preference could result in multiple types of music given the shape, center, and spread of the data. In the following excerpt, the instructor asked students to share problem-solving actions about a problem they had worked on the previous day. The problem read:

This year, the school band decided to poll all 330 middle school students about their favorite kind of music. The kind of music that is liked by more than 20% of the students will be played at the spring concert. Forty-two students liked country music, 110 preferred pop music, 13 voted for rap, 127 said music from TV shows like High School Musical, and 38 students tend to listen to rock. The band director wants a meaningful data display, an answer to her question, and for you to describe the (1) spread of the data and (2) whether there are any outliers.

T: What did you do [to solve the problem]?

S1: Highlighted and underlined and drew a little thing [bar graph]. There were five categories that people could vote for, so I split it up into categories and all the students and then that's. Since I knew how many students I had, then I started, I set that up and it easily laid it out for me so that I could start solving the problem.

...

T: Did anyone do it differently?

S2: I was going to do something different but then it didn't work out. What I was going to do is first make a bar graph so I can compare how many people like what. Do you want me to draw a bar graph on my paper so you can see it?

T: Yeah, why don't you do that and then we can come back to you.

The instructor routinely encouraged students to share their ideas so the

entire class might further explore and critique them. They were given data, asked to analyze it, and to determine the type of music that should be played at the next school dance, which might include multiple types if the data supported that conclusion. After individual think time followed by time for sharing ideas in pairs, the class reconvened to talk about their problem solving. The following excerpt starts after one student shared his thinking about the problem and had returned to his seat.

T: Did anyone do it [the problem] differently?

S1: [Walks to document camera and slides paper underneath it.] What I did was, if you see like this (points to projection) I put three hundred and thirty up there and then I put them up into those [genres of music].

S2: Forty-two divided by 330 because that's how you get percent, and I did that for each of them. And then for each of the percents, I either had to round up or round down. Like this one, you had to round up because that's a seven and seven is bigger than five, so you round up. ... I found which ones were higher than twenty percent because it said on this side. The kind of music that is liked by more than twenty percent will be played. But I had a problem because ... there were two that were bigger than twenty and I didn't really get that. ... the outliers are 127 kids that like TV music and 13 kids that like rap. That's what I did.

T: Is it possible that more than one type of music might be liked by more than 20% of students? [Several students nod affirmatively.] What do you think about her idea that she shared? [Several students nod affirmatively followed by a pause for the student to return to her seat.] Did anyone else do it differently?

S3: Well I think, you figure out how to divide 330 into 100, how to make 330 one hundred by dividing and then take that number, divide each of the categories and then you have a percent.

S4: You take 330 and then figure out what you need to do. What you need to divide by to make it 100 and then you take that number and divide each of the number of kids by that number and you get a percent because it's out of 100.

This discussion is evidence of how students decontextualized the information from the problem, manipulated the quantities to answer the question, and wrestled with contextualizing the result as it related to the question.

During these lessons, the instructor frequently asked whether students had questions about the student's presentation, which usually resulted in a couple of student-initiated questions. Some asked for assistance (e.g., "Can you explain it again?") whereas others posed more probing questions (e.g., "Why did you do it that way?"). After one presentation concluded, students

were asked to offer another mathematical model or strategy related to the problem. At least one student presented a viable model or alternate strategy for each problem. Students and the instructor ended the discussion when students' questions were answered and the classroom community believed (a) the problem was solved and (b) at least two distinct approaches (i.e., different representations, procedures, or combination of both) had been shared. Thus, some instructional decisions (e.g., ending a discussion) were jointly made by both the teacher and students. Following this discussion, students agreed on what music should be played at the dance. The whole-class discussion usually took 20-25 minutes.

Finally the instructor synthesized students' work and offered a summary of concepts, models, and analytic techniques that came up while solving the problem. Mathematics topics were often the focus of the synthesis. For example, the instructor shared how the range of a set of numbers provided different information about a data set than measures of central tendency. Students occasionally added to the instructor's synthesis and offered what they learned from solving the problem. The individual, small-group work time and whole-class discussion usually lasted 40-65 minutes.

Closure. During the last five minutes of class, concluding activities such as exit slips, reexamining introductory tasks, and writing summaries of the lesson were completed. One closure activity, from a class meeting different from the previous two discussed earlier, required students to share their ideas about statistical terms. The instructor asked students to define the word "cluster" after an earlier lesson. Many students commented that they were uncertain how best to characterize clusters of data.

S1: Clusters, it's a group or a pack.

S2: A bunch of things together.

S3: Say like in a number line, there's a bunch of numbers around five, six, and seven and there's like nothing for awhile, and then there's 19, 20, 21 it's just like, there's a lot of stuff in one area.

T: Okay, so, what do you think is an example of cluster?

S1: Well, to me, clusters are like groups. I would think of it like a pack of wolves. Like six and two packs of wolves, how many wolves are there?

S4: I'm sorry. I kind of disagree with your definition of it because this is talking about data and so groups can also mean like say there's a cluster of people who drive to school that are 18 or 19 and then there's less around 16.

As evident in this example, students shared and challenged each others' ideas. Materials such as summaries stayed in students' notebooks whereas exit slips were handed to the instructor as students left the room.

Problem-Solving and Unit Test Performance

A repeated measures t-test was conducted to answer the first research question: What is the impact of the intervention on students' performance on a test of word problems? Students in the intervention group performed better on the problem-solving posttest than the pretest, $t(17) = 2.65, p = .02, d = .48$, whereas their peers in the comparison group did not improve, $t(39) = 0.52, p = .61$ (see Table 6).

Table 6. Group means and standard deviations related to problem-solving performance and unit test performance

Factor	Intervention Group ^a		Comparison Group ^b	
	Mean	Std. Dev.	Mean	Std. Dev.
Problem-solving performance.				
Pretest	2.22	1.17	1.66	1.51
Posttest	2.83	1.34	1.73	1.28
Unit Test performance	17.11	3.69	19.88	3.07

^a N = 18; ^b N = 40

A one-way ANOVA was used to investigate the second question: Does performance on a test of word problems differ between students from the intervention and comparison groups? There was no significant difference between the intervention and comparison groups' pretest problem-solving performance, $F(1, 56) = 2.01, p = .16$. ANCOVA was employed to examine the relationship between posttest problem-solving performance and group status while holding pretest problem-solving performance constant. Students in the intervention group performed better than their comparison group peers ($M_{int} = 2.83, SD_{int} = 1.34; M_{com} = 1.73, SD_{com} = 1.28; F(1, 55) = 77.84, p \leq .005, d = .84$) (see Table 7).

Table 7. Problem-solving performance results

Variable ^a	SS	df	MS	F	partial η^2
Intercept	15.78	1	15.78	22.19*	0.29
Intervention Status	55.36	1	55.36	77.84*	0.59
Pretest Performance	5.98	1	5.98	8.41*	0.13
Residual	39.11	55	0.71		

* $p \leq .005$ ^a N = 58

Intervention status was uniquely associated with 13% of the total variance in posttest performance. Pretest performance and intervention status explained 72% of variance in students' performance on the posttest.

Finally, students' performance on the unit test was examined to answer the third research question: Does performance on a teacher-constructed unit test following TTPS instruction differ between students from the intervention and comparison groups? Mathematics FCAT scores were used as a covariate in the relationship between intervention status and unit test performance. Initial

ANCOVA results indicated that the covariate was not significantly related to posttest scores; therefore a one-way ANOVA was performed. There was a significant difference in the groups' mean scores on the unit test favoring the comparison group, $F(1, 55) = 8.27, p \leq .005, d = .79$ (see Table 8).

Table 8. *Unit test performance results*

Variable ^a	SS	df	MS	F	partial η^2
Intercept	16774.74	1	16774.74	1563.39*	0.97
Intervention Status	88.71	1	88.71	8.27*	0.13
Residual	590.14	55	10.73		

* $p < .005$ ^a N = 57

The comparison group had a higher average score than the intervention group ($M_{com} = 19.79, SD_{com} = 3.07; M_{int} = 17.11, SD_{int} = 3.69$). Similar to the problem-solving results, intervention status was uniquely associated with 13% of the variance in students' content knowledge.

Discussion

Findings from this study were both consistent and inconsistent with prior problem-solving research and offer information about the impact of TTPS as part of typical daily instruction during one sixth-grade mathematics unit. Intervention participants successfully answered more problems on the posttest than the pretest whereas their peers did not, much like the results from prior problem-solving explorations. The intervention group showed better problem-solving performance than the comparison group after one month of the instructional intervention. This finding is consistent with all of the past research on problem-solving instruction (e.g., Charles & Lester, 1984; Verschaffel & De Corte, 1997). Problem-solving test items were not explicitly constructed on the topics explored during the study. These findings provided evidence that TTPS supported students' problem-solving performance, regardless of the content embedded within problem-solving measures. Students showed some ability to transfer their experiences from the intervention period to problem-solving items that required students to draw on other mathematical concepts and procedures. One key finding of this study was that enacting TTPS instruction on a daily basis in a fashion described here led to improved problem-solving outcomes even after a short time period.

A second key finding is that the TTPS approach enacted in the present study did not help students respond correctly to unit-specific test items as much as typical teacher-led explicit instruction delivered by the comparison teacher. This is inconsistent with research on problem-solving instruction. Students experiencing problem-solving instruction tend to outperform their

peers experiencing explicit instruction (Sigurdson et al., 1994; Verschaffel et al., 1999). We take up this inconsistency more critically here through examination of two possible factors with an aim to stimulate thinking about the role of mathematics teaching, mathematical problem solving, mathematics content learning, and their interactions.

Realistic Tasks

The rich tasks in the intervention classroom provided a context for students to discuss mathematics content and procedures and engage in problem solving. The use of open, complex, and realistic word problems during instruction may foster cognitive links between students' prior knowledge (e.g., their mathematical knowledge and knowledge gained from experiences in the community; Boaler, 2002; Boaler & Staples, 2008; Palm, 2008). A word problem is realistic if its elements account for conditions in and out-of-school settings (Palm, 2006). Problems about local weather, pizza prices from local restaurants, and movie watching habits of local households provided a context for students to use their real-world knowledge in conjunction with their mathematics knowledge. Many (e.g., Boaler, 1993, 2002; Boaler & Staples, 2008; Matney, Jackson, & Bostic, 2013; Palm, 2008) have argued that realistic problems encourage children to draw on their knowledge from nonacademic situations, which may help them solve problems using novel approaches. On the other hand, some suggest that socioeconomic class strongly influences how students solve problems, which may limit their performance on realistic problems (e.g., Cooper & Dunne, 2000). In one study, students from lower socioeconomic households did not perform as well as their middle socioeconomic peers on items drawing upon realistic contexts (Cooper & Dunne, 2000). We cannot provide support for either argument about the influence of realistic problems due to low statistical power but we acknowledge that what may be realistic to one individual or group may not be realistic to another.

Pedagogy and Mathematics Learning

This study investigated students' performance on unit-specific mathematics items after experiencing TTPS. The average unit test score from the comparison group was approximately two points higher than the intervention group and we explore some potential explanations here.

First, it is possible that explicit instruction focused on learning procedures may have been a critical element linked to students' unit-test performance. Mathematics procedures were not made explicit during instruction within the intervention classrooms. Intervention students might have needed assistance abstracting mathematics procedures from the problem-solving experience and time spent practicing them, which may be a crucial element for supporting students' academic growth when employing TTPS. Davis (1992) suggests that

teachers “should start with problems or tasks, and as a result of working on these problems... a residue of mathematics... is what you have left over after you have worked on problems” (p. 237). The mathematical residue includes the procedures specific to solving a particular task, which may not have been explored adequately in the study’s lessons. Students need opportunities to practice employing procedures to solve mathematics tasks. Thus, students’ procedural knowledge growth might not have been sufficiently supported through the TTPS instructional approach enacted in this study, which might account for the differences in unit test scores.

Second and related to the first issue, students may be used to teacher-led explicit instruction and be unfamiliar with abstracting procedures from a problem. This has been documented in prior research (Arbaugh, Lannin, Jones, & Park-Rogers, 2006; Henningsen & Stein, 1997). Students in the intervention classroom might have needed more time to acclimate to this instructional approach. A third explanation is comparison students may have also been more prepared than their intervention peers for the types of questions found on the unit test. That is, the comparison teacher routinely administered tasks associated with the textbook materials. Daily assessment and instruction in the intervention classroom involved complex, realistic, and open-ended word problems, which were not found on the unit test. Intervention students’ lower scores on the unit test compared to their peers may be influenced by a misalignment between daily mathematics instruction and that measure. A similar argument might be constructed with the problem-solving performance differences and intervention students’ day-to-day engagement in problem-solving tasks. Future research could provide more valid evidence about students’ mathematics knowledge if both groups completed a unit test with both exercises and complex, realistic, and open-ended word problems.

Problem-solving Instruction

This study sheds light on a relevant instructional question: How might teachers teach mathematics content within problem-solving contexts? The results extend prior problem-solving studies by demonstrating that enacting TTPS on a daily basis within a supportive learning environment promoted better problem-solving performance than teacher-led explicit instruction. This study is further evidence that blending problem solving and mathematics instruction to achieve positive problem-solving and content knowledge outcomes is difficult. This statement, by itself, is not necessarily novel to the mathematics education field; however, the way the problem-solving instruction was conducted within the classroom was different from prior studies. Previous studies discussed earlier used a TTPS instructional approach sporadically (e.g., 20 times over four months) whereas the present investigation sought to examine students’ outcomes after employing TTPS everyday. Hence the present findings extend the mathematics education

field's knowledge base regarding students' outcomes from problem-solving instruction.

Students need instructional time to develop greater procedural fluency through exercises just as they need problem-solving experiences. This instantiation of problem-solving instruction, which did not allow students as much practice with procedures, may have limited their development of procedural fluency. Students needed time to practice procedures after using them within the problem-solving context. We conclude that teachers must make procedures (and practice with them) an explicit part of problem-solving instruction.

Limitations and Future Directions

This quasi-experimental mixed-methods study had limitations that impact the findings' generalizability. One limitation was that the sample size affected the statistical power of this study. A second limitation was evidence related to permanence of the intervention outcomes. This limitation arises in much of the past research on problem-solving instruction (e.g., Ridlon, 2009; Sigurdson et al., 1999; Verschaffel et al., 1999) and ought to be explored in a systematic fashion. A third limitation of this teaching experiment was an inability to randomly assign individual participants to each group. Statistical analyses suggested that the groups were similar in many aspects albeit this does not meet the randomized control assignment standards. The goals of this exploratory study were met; however, future investigations ought to draw on more students, sections, and teachers in order to explore the role of demographic variables in students' outcomes.

With two comparison classrooms and one intervention classroom, it was not possible through quantitative analyses to separate the effects of the instructor, intervention, and classroom. Future researchers might consider two classroom teachers conducting instruction in two classrooms in order to better separate the classroom and intervention effects and gain greater ecological validity evidence. A related question arises from the results: Might differences in instructors' content, pedagogical, and/or pedagogical content knowledge account for students' varied performance on the tests? Prior research suggests these bodies of knowledge likely impacted students' outcomes (e.g., Ball, Thames, & Phelps, 2008; Carpenter, Fennema, Peterson, & Carey, 1988). The instructor is a critical aspect of the intervention and there were some differences and similarities between instructors (see Table 9).

The instructor of the intervention group was a mathematics education doctoral student who taught elementary (i.e., grades K-6) mathematics methods courses over multiple semesters and previously taught in the middle grades (i.e., grades 6-8). The classroom teacher held a master of education degree and had been teaching middle grades mathematics at the school for approximately five years. The differences in the instructor and

Table 9. *Instructor and comparison teacher differences*

Factor	Instructor	Comparison Teacher
Years teaching middle school mathematics	3	5
Mathematics Coursework		
Undergraduate hours	41	3
Graduate hours	12	0
Mathematics Education Coursework		
Undergraduate hours	3	3
Graduate hours	12	0

teacher’s mathematical and pedagogical content backgrounds may have influenced the format and content of the instruction. This uncertainty could be resolved through a broader examination of the intervention across sections and instructors.

Summary

Enacting TTPS on a consistent basis during one mathematics unit was linked to both positive and negative outcomes for sixth-grade students. This study described an instantiation of TTPS during one unit and provided evidence that TTPS positively impacted students’ problem solving. TTPS as enacted in this study, which placed little emphasis on solving exercises and developing procedural fluency, might not support students’ mathematics content knowledge as measured by a unit test. The findings from this study suggest that TTPS should be *supplemented* with teacher-led explicit instruction rather than *replace* it as done in the present investigation. This conclusion is not contradictory to our results. Prior studies supplemented explicit (or otherwise) instruction with TTPS; results indicated students had better problem-solving and achievement outcomes than peers experiencing explicit instruction only. Further research is needed to explore other ways to implement TTPS, frequency of implementation (e.g., daily versus sporadically), and students’ outcomes from those experiences.

This exploratory teaching experiment characterized one way that TTPS might occur. Prior investigations provided guidance for this instructional intervention but TTPS had not been delivered on a regular basis during typical classroom duration and did not draw on state or national standards. The present study provides new evidence regarding the effects of TTPS and a description of TTPS instruction. In response to the first research question, intervention participants became better problem solvers as a result of daily TTPS. Intervention students had significantly better problem-solving performance than their peers in the comparison group after the intervention period. Finally, the comparison group did better than the intervention group on the unit test, responding to the third research question. This study provided

evidence about students' problem solving and content knowledge following daily TTPS instruction, which future researchers might explore in the era of accountability and CCSSM. For too long, problem solving has been treated "as an isolated topic akin to algebra or geometry. We need better integration of problem solving within all topic areas across the mathematics curriculum" (English & Sriraman, 2010, p. 267-268). If a goal of mathematics instruction is to develop competent problem solvers who are able to solve realistic problems, then teachers might consider *supplementing* their daily instruction with some form of TTPS to improve students' problem-solving performance.

Acknowledgements

The authors wish to thank Lieven Verschaffel and Helen Doerr for their feedback on this manuscript, as well as Toni Sondergeld for her assistance with RASCH analysis.

References

- Arbaugh, F., Lannin, J., Jones, D., & Park-Rogers, M. (2006). Examining instructional practices in Core-plus lessons: Implications for professional development. *Journal of Mathematics Teacher Education, 9*, 517-550.
- Australian Curriculum, Assessment, and Reporting Authority. (2014). *Mathematics F-10 Curriculum*. Retrieved from <http://www.australiancurriculum.edu.au/mathematics/Rationale>
- Ary, D., Cheser-Jacobs, L., Sorenson, C., & Razavieh, A. (2009). *Introduction to research in education* (8th ed.). Belmont, CA: Wadsworth.
- Ball, D., Ferrini-Mundy, J., Kilpatrick, J., Milgram, J., Schmid, W., & Schaar, R. (2005). Reaching for common ground in K-12 mathematics education. *American Mathematical Society, 52*, 1055-1058.
- Ball, D., Thames, M., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*, 389-407.
- Boaler, J. (1993). The role of contexts in the mathematics classroom: Do they make mathematics more "real"? *For the Learning of Mathematics, 13*, 12-17.
- Boaler, J. (2002). Learning from teaching: Exploring the relationship between reform curriculum and equity. *Journal for Research in Mathematics Education, 33*, 239-258.
- Boaler, J., & Staples, M. (2008). Creating mathematics futures through an equitable teaching approach: The case of Railside School. *Teachers College Record, 110*, 608-645.
- Bostic, J. (December 2012/January 2013). Model-eliciting activities for

- teaching mathematics. *Mathematics Teaching in the Middle School*, 18, 262-266.
- Bostic, J., Pape, S., & Jacobbe, T. (2011, October). *Validating two problem-solving instruments for use with sixth-grade students*. In L. Wiest & T. Lamberg (Eds.), *Proceedings of the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, pp. 756-763. Reno, NV: University of Nevada - Reno. Retrieved from <http://www.pmena.org/html/proceedings.html>
- Carpenter, T., Fennema, E., Peterson, P., & Carey, D. (1988). Teachers' pedagogical content knowledge of students' problem solving in elementary arithmetic. *Journal for Research in Mathematics Education*, 19, 385-401.
- Charles, R., & Lester, Jr., F. (1984). An evaluation of a process-oriented instructional program in mathematical problem solving in grades 5 and 7. *Journal for Research in Mathematics Education*, 15, 15-34.
- Cooper, B., & Dunne, M. (2000). *Assessing children's mathematical knowledge: Social class, sex, and problem-solving*. Philadelphia, PA: Open University Press.
- Cresswell, J. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th ed.)*. Boston: Pearson.
- Davis, R. (1992). Understanding understanding. *Journal of Mathematical Behavior*, 11, 225-241.
- Drouin, M., Horner, S., & Sondergeld, T. (2012). Alphabet knowledge in preschool: A rasch model analysis. *Early Childhood Research Quarterly*, 27, 543-554.
- Duncan, P., Bode, R., Lai, S., & Perera, S. (2003). Rasch analysis of a new stroke-specific outcome scale: The stroke impact scale. *Archives of Physical Medicine and Rehabilitation*, 84, 950-963.
- English, L., & Sriraman, B. (2010) Problem solving for the 21st century. In L. English & B. Sriraman (Eds.) *Theories of mathematics education: Seeking new frontiers* (pp. 263-290). Berlin, Germany: Springer. Florida Department of Education. (2007). *Florida mathematics standards*. Retrieved from <http://www.floridastandards.org/downloads.aspx>
- Henningsen, M., & Stein, M. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, 28, 524-549.
- Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H., Olivier, A., & Human, P. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.
- Hiebert, J., Carpenter, T., Fennema, E., Fuson, K., Human, P., Murray, ..., Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case for mathematics. *Educational Researcher*, 25(4), 12-21.

- Hiebert, J., & Wearne, D. (2003). Developing understanding through problem solving. In H. Schoen (Ed.), *Teaching mathematics through problem solving* (pp. 3-14). Reston, VA: National Council of Teachers of Mathematics.
- James, L., Demaree, R., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 68, 85-98.
- Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.
- Lambdin, D. (2003). Benefits of teaching through problem solving. In F. Lester, Jr., & R. Charles (Eds.), *Teaching mathematics through problem solving: Prekindergarten-grade 6* (pp. 3-13). Reston, VA: National Council of Teachers of Mathematics.
- Larson, R., & Boswell, L. (2010). *Big ideas: Math 6*. Erie, PA: Big Ideas Learning.
- Lesh, R., & Zawojewski, J. (2007). Problem solving and modeling. In F. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 763-804). Charlotte, NC: Information Age Publishing.
- Lester, Jr., F. (1994). Musings about mathematical problem-solving research: 1970-1994. *Journal for Research in Mathematics Education*, 25, 660-675.
- Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 7, p. 328. Retrieved from <http://www.rasch.org/rmt/rmt74m.htm>
- Linacre, J. M. (2002). *What do infit and outfit, measure and standardized mean?* *Research Measurement Transactions*, 16, 878.
- Linacre, J. M. (2006). *Winsteps® RASCH measurement (version 3.621)* [computer software and manual]. Beaverton, OR: Winsteps.com. Retrieved from Winsteps.com.
- Matney, G., Jackson, J., & Bostic, J. (2013). Connecting instruction, minute contextual experiences, and a realistic assessment of proportional reasoning. *Investigations in Mathematics Learning*, 6, 41-68.
- Ministry of Education, Culture, Sports, Science and Technology. (2007). *Arithmetic*. Retrieved from http://www.mext.go.jp/component/english/_icsFiles/afildfile/2011/03/17/1303755_004.pdf
- National Council of Teachers of Mathematics. (1980). *An agenda for action: Recommendations for school mathematics of the 1980s*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (2006). *Curriculum focal*

- points for Prekindergarten through grade 8 mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (2009). *Focus in high school mathematics: Reasoning and sense making*. Reston, VA: Author.
- National Governor's Association and Council of Chief State School Officers. (2010). *Common core standards for mathematics*. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf
- Palm, T. (2006). Words problems as simulations of real-world situations: A proposed framework. *For the Learning of Mathematics*, 26(1), 42-47.
- Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics*, 67, 37-58
- Ridlon, C. (2009). Learning mathematics via a problem-centered approach: A two-year study. *Mathematical Thinking and Learning*, 11, 188-225.
- Schoenfeld, A. (2011). *How we think: A theory of goal-oriented decision making and its educational applications*. New York: Routledge.
- Schroeder, T., & Lester, Jr., F. (1989). Developing understanding in mathematics via problem solving. In P. Trafton (Ed.), *New directions for elementary school mathematics, 1989 Yearbook of the National Council of Teachers of Mathematics* (pp. 31-42). Reston, VA: National Council of Teachers of Mathematics.
- Sigurdson, S., Olson, A., & Mason, R. (1994). Problem solving and mathematics learning. *Journal of Mathematical Behavior*, 13, 361-388.
- Smith, M., & Stein, M. (2011). Five practices for orchestrating productive mathematics discussions. Reston, VA: National Council of Teachers of Mathematics.
- Stein, M., & Smith, M. (1998). Mathematical tasks as a framework for reflection: From research to practice. *Mathematics Teaching in the Middle School*, 3, 268-275.
- Verschaffel, L., & De Corte, E. (1997). Teaching realistic mathematical modeling in the elementary school: A teaching experiment with fifth graders. *Journal for Research in Mathematics Education*, 28, 577-601.
- Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999). Learning to solve mathematical application problems: A design experiment with fifth graders. *Mathematical Thinking and Learning*, 1, 195-229.
- Verschaffel, L., Van Dooren, W., Greer, B., & Mukhopadhyah, S. (2010). Reconceptualising word problems as exercises in mathematical modeling. *Journal für Mathematik-Didaktik*, 21, 9-29.

Appendix A

1) Ruth is planning to serve ice cream sundaes to guests at her birthday party. She purchased 3 flavors of ice cream: vanilla, chocolate, and strawberry, 2 different sauces: chocolate and caramel, and 4 different toppings: bananas, nuts, sprinkles, and whipped cream. How many different types of sundaes can be made if every guest selects only one ice cream flavor, one type of sauce, and one topping?

2) A group of 150 tourists were waiting for a shuttle to take them from a parking lot to a theme park's entrance. The only way they could reach the park's entrance was by taking this shuttle. The shuttle can carry 18 tourists at a time. After one hour, everyone in the group of 150 tourists reached the theme park's entrance. What is the fewest number of times that the shuttle picked tourists up from the parking lot?

3) Aunt Marie purchased 80 Silly Bandz for her two nephews Elijah and Jordan. She gave Elijah 10 more Silly Bandz than Jordan. How many Silly Bandz did Elijah and Jordan each receive?

4) A family is planning a camping trip to a national park and receives the following information about the costs per day:

Camping Fee	
Children 12 years and younger	\$3.00 per day
All others	\$7.00 per day
Parking for trailer	\$9.00 per day
Use of common areas	\$1.50 per person per day

The family will camp for 10 days and need to park their trailer each day. The family consists of 4 people including a father, mother, 8 year-old child, and a 14 year-old child. Each person will need to use the common areas on a daily basis. How much will they pay for their camping trip?

5) Maria wanted a bicycle so she started saving all of her money. For every \$6.00 that Maria saved, her mother gave her \$2.00. Maria had \$56.00 after three months. How much money did Maria's mother give her?

Appendix B

Southernville Pizza

Directions: Use your knowledge of ratios, rates, unit rates, data representations, and data analysis to answer the questions below. Please show all of your work for every problem-solving step. Create a math model and use a strategy to find the result for each question. Carry out your work here and use the back of the paper, if needed. Answer all questions in complete sentences that fully justify and explain your solution.

The city of Southernville has many places to purchase a pizza. Jeremy decides to create a website to provide residents with information that may help them decide where to purchase their pizza. The following data provide the cost of a cheese pizza, a pepperoni pizza, a large pizza with five toppings, the diameter of a large pizza, and the number of slices on a large pizza:

Pizza Restaurant	# of Slices on Large pizza	Diameter of Large Pizza (in.)	Cost of Large Cheese Pizza (dollars)	Cost of Large Pepperoni Pizza (dollars)	Cost of Large Pizza with 5 Toppings (dollars)
Pizza Hut	8	14	10.00	10.00	10.00
Papa Johns	8	14	8.99	9.99	12.99
Domino's	8	14	9.99	7.99	15.06
Five Star	8	14	8.99	10.49	12.99
Leonardo's	8	14	8.75	10.95	16.50
Hungry Howie's	8	14	10.55	12.95	16.05
Pizza Vito	8	14	10.95	12.70	19.95

Q1: Create a data representation that Jeremy might display on his website to help customers decide on what pizza to buy from a restaurant.

Q2: Write a letter describing the best value for a pizza that your family might be interested in purchasing. Write in a way that a 6th grade student might understand.

*****Check your work with one other person or another group of people. If they have something different, write it in pen near your answer because we will discuss them later.*****

Appendix C

A guide to the six stages of problem solving

1. Reading the problem.
 - a. Did you read the entire problem?
 - b. Were there any words that you need help understanding?
 - c. Do you understand what you are supposed to find?
2. Describing the situation
 - a. What is happening in this problem?
 - b. Can you represent the situation presented in the problem?
3. Creating a mathematical model
 - a. What information is necessary to solve the problem?
 - b. What information is unnecessary to solve the problem?
 - c. Think about whether this problem is similar to others you have seen before.
 - d. Is there more than one way to begin solving this problem?
4. Using a strategy and finding the result.
 - a. Think about some possible strategies and choose one that will work with what you created in the previous stage.
 - b. Look at your work thus far. Did you make any mistakes with your arithmetic or carrying out the strategy?
 - c. Does your result make sense when you look at your mathematical model?
5. Interpreting your result
 - a. What are the units for your result?
 - b. Does your result answer the original question?
 - c. Does your result fit with your situation? Is it a realistic answer?
6. Reporting your answer
 - a. Did you write a sentence that clearly answers the question with the final solution?

** Is there another strategy that might answer the problem? Does your strategy use different steps to calculate the result? **
--