Journal of Mathematical Behavior 65 (2022) 100933

Contents lists available at ScienceDirect

Journal of Mathematical Behavior

journal homepage: www.elsevier.com/locate/jmathb

Fourth-grade students' sensemaking during multi-step problem solving

Gabriel Matney [a,*], Jonathan D. Bostic [a], Miranda Fox [a], Tiara Hicks [a], Toni May [b], Greg Stone [c]

[a] Bowling Green State University, 529 Education Building, Bowling Green, OH, 43403, United States
[b] Drexel University, 3401 Market St., 3rd Floor, Philadelphia, PA, 19104, United States
[c] Metriks, MetriKs Amérique, 5840 Summit St., Sylvania, OH, 43560, United States

ARTICLE INFO

Keywords:
Sensemaking
Problem solving
Word problems
Strategy
Elementary students

ABSTRACT

The purpose of this study was to investigate fourth-grade students' sensemaking of a word problem. Sensemaking occurs when students connect their understanding of situations with existing knowledge. We investigated students' sensemaking through inductive task analysis of their strategies and solutions to a problem that involved determining the difference between two quantities and number of groups within the task. This analysis provided useful themes about students' strategy use across the sensemaking domains. Students exhibited three levels of sensemaking and many different strategies for solving the problem. Some strategies were more helpful to students in achieving a correct result to the problem. Findings suggest that sensemaking about problems involving differences and number of groups is difficult for many fourth-grade students. Among students who did make sense of the problem, those who used efficient strategies with more familiar operations tended to do better than those who used less efficient strategies or algorithms.

1. Introduction

For more than 40-years, the National Council of Teachers of Mathematics has advocated for mathematics instruction to focus on problem solving (NCTM; 1980, 1989, 2000, 2014, 2020). A common means to engage students in mathematical problem-solving experiences is through word problems (Bostic et al., 2016; Levingston et al., 2009; Verschaffel et al., 2007a). Word problems are an important part of elementary mathematics instruction and assessment (Bostic et al., 2016; Verschaffel et al., 2007a). Understanding students' sensemaking during word problem solving is a key to fostering students' success while problem solving (Pape, 2004). Problem solving has been a topic of research for many decades and continues to be an object of inquiry (e.g., Matney et al., 2013; Sahin et al., 2020; Schindler & Bakker, 2020; Verschaffel et al., 2009; Yee & Bostic, 2014).

The problem solving study we present here involved students from the United States of America (USA), which has the added context of the Common Core State Standards for Mathematics (CCSSM; CCSSI, 2010). In 2010, the majority of states moved to adopt the CCSM, which were based in research, focused, coherent, and aligned with international benchmarks for the learning of mathematics (Carmichael et al., 2010; Schmidt & Houang, 2012). Along with these qualities, and directly related to study in this paper, the CCSSM

* Corresponding author.
E-mail addresses: gmatney@bgsu.edu (G. Matney), bosticj@bgsu.edu (J.D. Bostic), foxmir@bgsu.edu (M. Fox), hickstr@bgsu.edu (T. Hicks), tas365@drexel.edu (T. May), gregorystone@metriks.com (G. Stone).

https://doi.org/10.1016/j.jmathb.2022.100933
Received 2 February 2021; Received in revised form 10 January 2022; Accepted 19 January 2022

established real-life problem solving as something students should be engaging in throughout their K-12 education (CCSSI, 2010). Furthermore, teachers across states were expected to promote students' mathematical proficiency by providing opportunities for students to "make sense of problems and persevere in solving them" (CCSSI, 2010, p. 6). Thus, we desired to inquire about students' sensemaking during problem-solving following the implementation of the CCSSM.

It is important for the field to continue to conduct problem-solving research with students, as these data can inform us about students' struggles and how mathematics instruction may be optimized. Through the investigation shared here, we sought to illuminate students' sensemaking of word problems, the strategic choices students made, and the levels of success those choices afforded them. We discuss implications of these findings for the mathematics education community.

2. Theoretical frameworks and related literature

This study is framed by notions of problem solving and sensemaking about word problems.

2.1. Problem solving and word problems

Broadly speaking, problem solving "is what you do when you don't know what to do" (Sowder, 1985, p. 141). From this perspective, problem solving is a key feature of different types of mathematical engagements, such as, realistic or real-life settings (Bostic et al., 2016; Matney et al., 2013; Verschaffel et al., 1999), word problems (Bostic et al., 2016; Palm, 2006, 2008; Reed, 1999), problems that involve discerning mathematical structure and relationships (CCSSI, 2010; Schoenfeld, 1992), application problems (Verschaffel et al., 1999), and open-ended tasks (Becker & Shimada, 1998; Cifarelli & Cai, 2005). Each kind of mathematical engagement brings its own affordances and has a place in mathematics education (Verschaffel et al., 2000; Verschaffel et al., 2020). Problem solving goes beyond the type of thinking needed to solve exercises (Mayer & Wittrock, 2006) and occurs when the task is a problem, not an exercise (Schoenfeld, 2011).

We utilize Verschaffel and colleagues' (2000) problem-solving framework to consider students' sensemaking of word problems. This six-stage framework of problem solving includes: (a) reading the problem, (b) creating a representation of the situation, (c) constructing a mathematical representation of the situation, (d) arriving at a result from employing a procedure on the representation, (e) interpreting the result in light of the situational representation [see (b)], and finally, (f) reporting the solution within the problem's context. This framework dovetails with the way Verschaffel et al. (1999) characterizes word problems as (a) open, (b) developmentally complex, and (c) realistic tasks for an individual. *Open* problems can be solved using multiple developmentally-appropriate strategies. Problems are *developmentally complex* for a student when they require productive thinking and are not readily solvable (Schoenfeld, 2011). Complex problems are distinct from exercises; exercises are tasks intended to promote efficiency with a known procedure (Kilpatrick et al., 2001). *Realistic* tasks may draw upon real-life experiences, experiential knowledge, and/or believable situations (Verschaffel et al., 1999).

Generally speaking, word problems are mathematical tasks presented in text, which contain real-life situational background information (Verschaffel et al., 2000). Arithmetic word problems necessitate the use of the operations of addition, subtraction, multiplication, and/or division. For our considerations here, the term 'word problem' constitutes a situational mathematics task presented in text format and necessitating arithmetic operations for its solution. These frameworks for problem solving and word problems guided our study's investigation. We were particularly interested in the mathematical strategies students enact as a result of their sensemaking of arithmetic word problems that have these attributes.

2.2. Sensemaking about multi-step word problems

Sensemaking is when students develop an understanding of a situation or context by connecting it with existing knowledge (NCTM, 2009, p. 4). Students who struggle to comprehend the relationship between textual elements representing the situation will consequently struggle to associate the correct mathematical operations (Nortvedt, 2010; Pape, 2004; Verschaffel et al., 2000). Hence, problem solving and sensemaking are connected in that "becoming a good mathematical problem solver—becoming a good thinker in any domain—may be as much a matter of acquiring the habits and dispositions of interpretation and sense-making as of acquiring any particular set of skills, strategies, or knowledge" (Resnick, 1988, p. 58).

Of particular interest for our study is the notion of students' ability to decontextualize, meaning, "to abstract a given situation and represent it symbolically and manipulate the representing symbols as if they have a life of their own, without necessarily attending to their referents" (CCSSI, 2010, p. 6). As students are sensemaking while problem solving, the ability to decontextualize enables them to connect the situation being presented in the task to the mathematical operation necessary for a strategy (Verschaffel et al., 1999). This decontextualization process includes mathematical strategy use, which we define as a mathematical pathway that an individual enacts while problem solving, and includes both representations and operational procedures (Goldin, 2002). Strategic choices in open problems are influenced by students' sensemaking of the context (Matney et al., 2013), their connection between the context and known applicable arithmetic calculations (Baroody & Dowker, 2003; Kilpatrick et al., 2001; Yee & Bostic, 2014), and their ability to execute those calculations (Torbeyns et al., 2006; Verschaffel et al., 2007b). A complementary sensemaking attribute is contextualization. Contextualization occurs during sensemaking as students' reference back to the problem context to check the continued relevance of their mathematical work (Koestler et al., 2013). Mathematically proficient students use both of these abilities while sensemaking about the quantities within the situation presented in a word problem (CCSSI, 2010).

Sensemaking occurs at many stages of problem solving (Verschaffel et al., 2009) and some scholars have focused on students' work between the situation and mathematical representations. For instance, Palm's (2008) qualitative study examining fifth-grade students' work indicated that students' engagement with realistic word problems increased the likelihood their problem solving ended with a correct solution to a problem. Similarly, Yee and Bostic (2014) conducted a qualitative study examining sixth-grade and high school students' word problem solving and drew a conclusion that more successful problem solvers were flexible with their mathematical representations during strategy use, often using multiple nonsymbolic representations, compared to peers who employed solely symbolic representations while problem solving. Taken collectively, the literature provides ideas about students' strategy use while problem solving, but few take a critical look at students' work to explore their mathematical sensemaking of word problems. Hence, this study aims to fill a needed gap within the problem-solving literature.

As part of students' sensemaking, they draw upon different representations of the situation within the problem – and its mathematical elements - to organize and externalize their thinking (Cai & Lester, 2005; Goldin, 2002). Misunderstandings while sensemaking can generate barriers for students to successfully problem solve (Yee & Bostic, 2014). In prior research it was demonstrated that students had a strong tendency to give 'unrealistic' solutions to word problems (Greer, 1997; Palm, 2008; Yoshida et al., 1997). More specific to the context of the study here, students solving arithmetic word problems in school settings tended to use one or more arithmetic operations involving the quantities in the problem without consideration of the context (Verschaffel et al., 2000). Development of sensemaking habits of mind help students develop autonomy, relying on their own reasoning and resources to be more persistent in problem solving (Mueller et al., 2011; Yackel & Cobb, 1996). Hence, in order for educators to help students become better problem solvers, it is important for them to know what students are, or are not, making sense of during problem solving.

2.2.1. *Making sense of the difference*

Some word problems necessitate making sense of a context that requires the difference between two or more quantities. Two common models for understanding approaches to subtraction are taking away and determining the difference (Selter et al., 2012). In the take away model, students invoke a strategy involving *taking away* the subtrahend amount from the minuend, as in $433-379 = 54$. Alternatively, students might notice the relatively close nature of the two numbers (379 and 433) and approach the difference with a strategy of starting with 379 and *determining the difference* up to 433. As an example, consider this case where an individual must make sense of a difference between quantities:

> Ashley and Jorge are both collecting unique pencils. Ashley wants to have as many pencils as Jorge. Jorge has 433 and Ashley has 379. How many more pencils does Ashley need so that Ashley's collection is the same size as Jorge's?

To solve this an individual needs to observe that a difference needs to be found. Procedural expectations for finding the difference in fourth-grade are found in the CCSSM when it says students should "fluently add and subtract multi-digit whole numbers using the standard algorithm" (CCSSI, 2010, p. 29). The standard algorithm in the USA operates from a take away perspective. However, there are many different operational procedures that young children might use to find the difference between two numbers (Caviola et al., 2018). Students may account for the difference by adding up (Torbeyns et al., 2018), by using a decomposition strategy, or by using a subtraction procedure other than the standard algorithm. A student's choice of procedure in the context of finding a difference during strategy use, such as adding up or a subtraction algorithm, goes beyond their comfort with a particular procedure (Smith-Chant & LeFevre, 2003). The choice may also be influenced by the problem's difficulty in terms of the magnitude of the numbers involved (Caviola et al., 2018) or the necessity of regrouping parts of an operand in order to subtract (Imbo et al., 2007), which relates to how students initially made sense of the problem.

Students often lack strategic flexibility (Torbeyns et al., 2006) and tend to use the same operational procedure regardless of the numbers (Selter, 2001). Selter (2001) conducted a qualitative study with 300 third-grade students, exploring how they solved addition and subtraction tasks that lacked context (e.g., $836-567=?$). Results indicated that "43 % of the students did all problems by means of the [standard] algorithm" (p.160). This included the task: $701-698$, which can be much more simply solved by recognizing the minuend and subtrahend are only three apart. Furthermore, about 75 % of the students used the same operational procedure for solving every subtraction problem, or about 83 % if you include students who used a second operational procedure on only one out of six problems. Our work extends from this study by utilizing a multi-step arithmetic word problem. With the CCSSM's emphasis on arithmetic strategies and properties of operations in the elementary grades (CCSSI, 2010), part of our investigation inquires about those strategies students are using to find the difference between quantities in word problems.

2.2.2. *Making sense of the number of groups*

Relative to addition, subtraction, and multiplication, research on multi-digit division has received limited attention (Hickendorff et al., 2018). Some word problems necessitate making sense of a context in which the students need to operate with groups of a particular quantity. "For equal-sized group situations, division can require finding the unknown number of groups (Quotative) or the unknown group size (Partitive)" (CCSSI, 2010, p. 21). Making sense of the number of groups would be required for the situation posed in the earlier task with Ashley and Jorge.

> Ashley has 54 fewer pencils as Jorge and wants to buy enough pencil packs so that Ashley's collection is as large as Jorge's collection. The pencil packs Ashley likes best come with three pencils per pack. What is the number of pencil packs Ashley needs to buy to have as many pencils as Jorge?

Making sense of this context necessitates observing the number of groups of 3 composed within 54. Thus, the problem context dictates a logic that the group size is known, and the number of groups is unknown. Similarly, in this study students were challenged with a multi-step word problem involving finding a number of groups, opening up an option for students to choose multi-digit division as part of their overall strategy. The CCSSM indicates that students should be able to solve division problems using "strategies based on place value, the properties of operations, and/or the relationship between multiplication and division" (CCSSI, 2010, p. 30).

Calculating the number of groups of a particular size is demanding for upper elementary students (Fagginger Auer et al., 2016; Hickendorff et al., 2009; Schulz & Leuders, 2018). Similar to making sense of the difference, students enact many different operational procedures to find the number of groups (Anghileri et al., 2002; Hickendorff, 2013; Hickendorff et al., 2010; Schulz & Leuders, 2018). Hickendorf et al. (2009) classify these operational procedures in two ways: number-based or digit-based. Number-based operational procedures allow students to safely underestimate and use simpler facts that they are more familiar with (Kilpatrick et al., 2001). These procedures include those such as repeated addition and/or subtraction, estimation using multiplication, compensation, and partial quotients. When students use repeated addition, they repeatedly add the same addend until reaching a desired sum. In the language of division, they add multiples of the divisor until they reach the dividend. Repeated subtraction then proceeds in the opposite direction as students repeatedly subtract multiples of the divisor until reaching a desired difference. In a more condensed approach, students may use multiplication strategies in order to determine the number of multiples that make up the dividend. Students who do this often start by multiplying the divisor by a number that may be close to the dividend and then proceed using an informed trial and error approach until they find the number that results in the dividend. In compensation, the dividend or the divisor may be rounded up or down to a multiple of the divisor. For example, when a student needs to know how many groups of 3 are in 54, they could estimate the number of groups as $3 \times 20 = 60$, since 60 is close to 54. Noting this is too large, the student may perform the calculation $60-54 = 6$, and thereby recognize the number of groups of 3 should be 18. This procedure requires students to account for the fact that the multiplier (in this case 20) represents the number of groups. Lastly, the partial quotients procedure is an approach to division in which the student decides how many groups of the divisor to subtract from the dividend. The number of groups of the divisor is kept track of and the process of subtracting multiples of the divisor continues until the entire dividend is comprised. The total number of groups of the divisor is then summed to find the quotient. The partial quotients method of division is generally a more accessible approach as it allows for students to use safe underestimations and find the quotient part by part (Kilpatrick et al., 2001).

Digit-based operational procedures, such as the standard division algorithm, rely on students' procedural fluency in accurately replicating a procedure. In order for digit-based procedures to be meaningful, students need to develop a conceptual understanding of why the process works, including an understanding of the base ten system (Kilpatrick et al., 2001). Kilpatrick et al. (2001) claim that the division algorithm provides a challenge for students because it requires them to find the exact number of copies of the divisor that can be taken from successive parts of the dividend. This method leaves little room for error. Just as with finding the difference, the procedures students use when finding the number of groups is largely based on the magnitude of the numbers, preexisting knowledge of mathematical facts, as well as the procedures that they have learned at home and at school (Kilpatrick et al., 2001).

2.3. The fair task

In this study, we investigate students' sensemaking of an arithmetic word problem called the Fair Task. The Fair Task is a non-routine, multi-step, open, complex, realistic, arithmetic word problem that addresses the CCSSM domain of Operations and Algebraic Thinking. More specifically, the task addresses elements of Standard 4.OA.3:

> Solve multi-step word problems posed with whole numbers and having whole-number answers using the four operations. Assess the reasonableness of answers using mental computation and estimation strategies including rounding (CCSSI, 2010, p. 29).

The Fair Task was designed for fourth-grade students using the frameworks for problem solving and word problems discussed earlier. The task comes from a series of problem-solving measures for grades 3–8 (i.e., Problem solving Measures or PSMs). The PSM series consists of mathematics word problems aligned with the grade level content of the Common Core State Standards for Mathematics (CCSSI, 2010). PSM items have been rigorously explored through multiple validation studies, including think alouds with students and expert panels with representatively sampled mathematics teachers, mathematicians, and mathematics educators (Bostic & Sondergeld, 2015; Bostic et al., 2017, 2018, 2019, 2020, 2021). The Fair Task is one of 15 items found on the PSM4 and is shown below.

> Josephine sold tickets to the fair. She collected a total of $1,302 from the tickets she sold. $630 came from the adult ticket sales. Each adult ticket costs $18. Each child ticket costs $14. How many child tickets did she sell?

The Fair Task incorporates multi-step thinking in the areas of making sense of the difference and making sense of the number of groups.

2.4. Three key observations of sensemaking in the fair task

In the Fair Task, students encounter a problem context involving the purchase of tickets for children and adults. They are given the cost of each type of ticket and total values for all child and adult tickets sold. Students must use this information to consider how many child tickets were sold. In solving this task, students need to determine which pieces of information are relevant and helpful. The $18 cost for adult tickets is not necessary to finding the solution of the number of child tickets sold. On the other hand, the numbers 1302, 630, and 14 are helpful for moving through the problem with three key sensemaking observations, one involving a difference, one involving groups of 14, and the other involving the connection between groups of 14 and the difference. We have numbered the key observations (KO) for the sake of convenience, not to prescribe an assumed order through which students must make sense of the problem.

KO1) Students need to observe that the difference between $1,302 and $630 is the dollar amount brought in by selling child tickets. This value is $672.

KO2) Students need to observe that each child ticket is $14 dollars. Hence, there is some number of groups of 14 that would represent the number of child tickets sold.

KO3) Students need to observe that the number of groups of 14 within the difference of 672 will give them the number of child tickets sold.

These three key observations demonstrate the constraints and affordances of the Fair Task. Although students' processes are constrained by the given conditions and one possible correct answer, decisions on how to account for each of the observations remains at the mathematical discretion of the problem solver, which invokes an important degree of strategic openness through which students may approach solving the task. When students make sense of KO1, they may choose to conceptualize the difference as 1302 *take away* 630 or *determine the difference* between the two numbers using another strategy. When students make sense of KO2, the context gives the group size of 14, making the division an unknown number of groups (Quotative) situation. However, students may employ a number of different strategies for finding the number of groups. We know from prior research (Quintero, 1983) that students struggle to make sense of multi-step problems involving ratio. Yet, little is known about students' sensemaking of multi-step problems involving combinations of operations, such as a difference and a number of groups. Further, less is known about the ways students approach multi-step arithmetic word problems. The research here works to fill that gap by describing students' sensemaking for a multi-step problem and evaluating the patterns of success in their strategic choices.

We wondered what evidence students might demonstrate for sensemaking about each key observation. The multiple elements of the Fair Task afford students both a challenge and an opportunity to make sense of the context and choose a solution pathway, strategically. The aims of this manuscript are (a) to describe the sensemaking evidence of fourth-grade students and (b) to investigate patterns among students' strategy use for success and connections to sensemaking. More specifically, we inquire about the following research questions (RQ).

RQ1: What sensemaking is evident from students' solutions to the Fair Task?

RQ2: What strategies do students use to solve the Fair Task?

RQ3: How do those strategies relate to student success on the Fair Task?

With a better understanding of students' sensemaking while problem solving and the degree of success that their mathematical strategies afford them, educators and scholars may be better able to make evidence-based decisions about instruction that promotes problem solving.

3. Method

3.1. Validity evidence for the fair task

Validation studies for the PSMs that involved the Fair Task adhered to rigorous expectations for developing high quality assessments as framed by the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], et al., 2014). Necessary and appropriate sources of validity evidence were gathered and analyzed for the Fair Task including test content, response processes, relations to other variables, internal structure, and consequences from testing/bias (AERA et al., 2014). An expert panel consisted of professionals including elementary teachers, school instructional leaders, and terminally degreed mathematics educators and mathematicians familiar with the content and expectations of the CCSSM (CCSSI, 2010). Think alouds were performed with purposefully and representatively selected students as well as whole-class think alouds (Bostic et al., 2021). The Fair Task was administered to students in multiple districts over several years, which were representatively and purposefully chosen for their different student populations, representing diversity of racial and ethnic identities as well as geographic location (i.e., rural, suburban, and urban). Psychometric analysis in prior validation studies used Rasch analysis (Rasch 1960/1980) to reveal that the Fair Task had a 0.43 logit value, reasonable infit (0.85) and outfit (.73) values, and positive point-biserial values. This logit value within the context of other PSM4 items led to the impression that this item was moderately difficult. The item discrimination and other psychometric findings were acceptable (see Bostic et al., 2019; Linacre, 2021). Finally, qualitative (e.g., interview data with respondents and expert panel) and quantitative findings (e.g., comparisons of different subgroups) indicated that this item had little bias impacting respondents' outcomes. Detailed information about the development of items in the PSMs, their associated validation process, and each facet may be found in related publications (i.e., Bostic & Sondergeld, 2015; Bostic et al., 2017, 2018, 2019, 2020, 2021). Robust validity evidence related to the Fair Task and PSM4 convey assurance that users can feel confident that score interpretations and use of the tasks with fourth-grade students are appropriately drawn.

3.2. Participants and setting

The 280 participants in this study were fourth-grade students from two school districts in a Midwest state located in the USA using the CCSSM as the basis for their curricular standards. District A is a rural district and District B is a suburban district. Demographic

Table 1
District Demographics.

District	K-12 Student Population	Female	Male	White	Non-White	Poverty	Disability
A	1,915	48.83%	51.17%	83.86%	16.14%	36.14%	15.30%
B	3,530	49.52%	50.48%	60.76%	39.24%	39.38%	13.06%

information for each district can be found in Table 1. The Fair Task was given to participants as part of the fourth-grade PSM (PSM4) in the last month of the 2018 academic year. Students worked on the PSM4 individually, in a quiet classroom setting monitored by the researchers and the students' classroom teacher. They were not permitted to use calculators and were encouraged to write, draw, and represent their ideas on the testing paper. Problems were administered on single-sided paper and students were permitted to use both sides of paper or additional paper. Teachers provided enough time for all students to complete the PSM4, which was typically 90−120 min. The Fair Task was the fifth problem on PSM4. All names of students used in this manuscript are pseudonyms.

3.3. Data and analysis

Participants completed the Fair Task and expressed their problem solving in writing on the PSM4. The written work on the Fair Task was scanned for ease of analysis. An inductive qualitative analysis (Hatch, 2002) was performed to find thematic (Creswell, 2011) patterns among the mathematical work of students to illuminate findings related to each research question. In what follows, we share our seven step process for the enacted inductive analysis.

1. Three researchers read the solutions of all 280 students. The frame for the analysis was evidence of student mathematical sensemaking of the problem related to the three key observations for the Fair Task. Researchers looked for sensemaking as evidenced by student work that demonstrated understanding of the Fair Task's context and connected it to some mathematical strategy for solving the problem at hand.
2. The three researchers reread the 280 students' solutions and made memos of salient domains.
3. Researchers collaboratively identified salient domains and gave them a code. Each of these domains consisted of particular groups of strategies.
4. The researchers reread the data and refined and revised salient domains, grouping like strategies and recording which students calculated correctly, or incorrectly, for their chosen strategy.
5. Each of the three researchers took a specific domain and reread all of the students' solutions to decide if the domains were supported by the data. Discrepancies were shared with the research team and discussed for consensus agreement. When appropriate, the discrepancies were analyzed for their merit within other groupings of strategies or considered as a new strategy. This completed the analysis for RQ1. The authors created written paragraphs and graphic maps to demonstrate each domain. These are further explicated and shared in the findings below.
6. Completed domains were analyzed, within and across, for patterns involving students' solution strategies for the Fair Task. When patterns among student strategies were found, focused analysis on the participants' work exhibiting those patterns was conducted.
7. For each pattern, the researchers created an outline expressing relationships within and among domains. This completed the analysis for RQ2 and RQ3. Data excerpts to support the elements of the outline were selected. A textual outline was made first, followed by a picture meant to capture the findings.

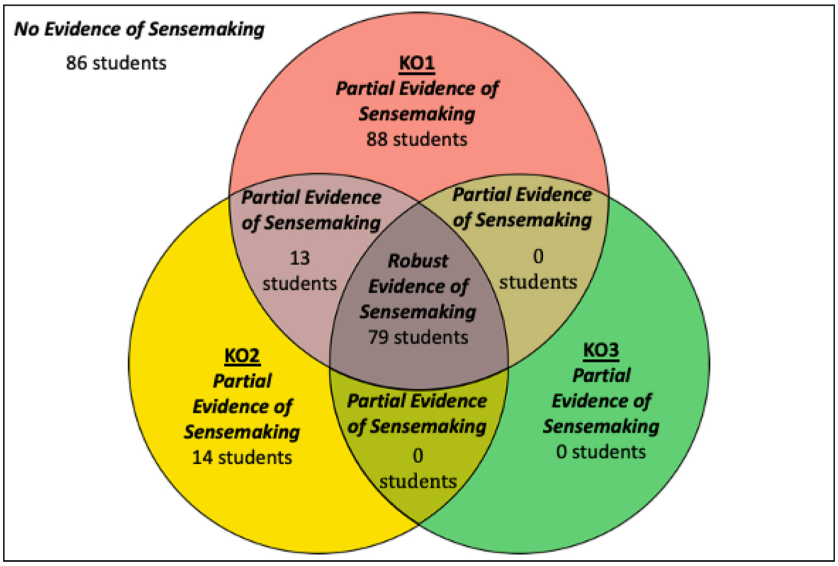

Fig. 1. Venn diagram showing the relationship between the three levels of sensemaking (Robust, Partial, and No) and the three key observations (KO) of the Fair Task. KO1-The difference between $1,302 and $630 is the dollar amount brought in by selling child tickets. This value is $672. KO2-There is a number of groups of 14 that gives the number of child tickets sold. KO3-The number of groups of 14 within the difference of 672 yields the child tickets sold. The number of students found within each level KO are given.

4. Findings

The study here investigates evidence of students' sensemaking (RQ1), descriptively maps out students' strategies in relation to their sensemaking (RQ2), and shares analysis of strategy success (RQ3). Overall, only 50 out of 280 (18 %) students derived a correct result. These 50 students showed the sensemaking capability, strategic competence, and computational fluency to correctly solve the Fair Task. We further parse out the aspects students made sense of in the Fair Task.

4.1. RQ 1: evidence of sensemaking

Inductive analysis revealed three qualitatively different levels of student sensemaking. These domains were (i) robust evidence of sensemaking, (ii) partial evidence of sensemaking, and (iii) no evidence of sensemaking. Students' sensemaking about the Fair Task was observed as they demonstrated a mathematical strategy related to one or more key observation(s). Concisely stated, the three key observations were KO1-The difference between $1,302 and $630 is the dollar amount brought in by selling child tickets. This value is $672. KO2-There is a number of groups of 14 that gives the number of child tickets sold. KO3-The number of groups of 14 within the difference of 672 yields the child tickets sold. Fig. 1 shows the connection between the levels of sensemaking, the KOs, and the number of students identified in each. The figure also conveys that students who exhibited a robust sensemaking are in the intersection of all three KOs, students who exhibited partial sensemaking are in the other component parts of the KOs, and students who exhibited no sensemaking fall outside of the three KOs. As students made sense of the problem, some of them also understood the *connection* between KO1 and KO2, thus demonstrating sensemaking of KO3. Therefore, sensemaking about KO3 only happened when accompanied by the first two.

4.1.1. Robust evidence of sensemaking

Robust evidence of sensemaking about the Fair Task indicated attention to all three key observations necessary to solve the problem. As seen in Fig. 2, 79 out of the 280 students (28 %) provided evidence that they made sense of KO1, KO2, and KO3 and enacted 14 unique mathematical strategies to derive an answer. Though these strategies varied, all 79 students used a viable strategy with 50 students arriving at the correct answer and 29 students making some form of minor miscalculation. This suggests that these students showed evidence of making sense of the difference, the number of groups, and that the number of groups within the appropriate difference yielded the desired solution.

Fig. 2. Mapping of strategies for students who demonstrated robust evidence of sensemaking in their solving of the Fair Task.

Fig. 3 offers four samples of student work evidencing robust sensemaking. Students showing robust evidence of sensemaking for KO1 all implemented the standard algorithm for subtraction to find the difference. Students' strategic choices varied the most in their attention to finding the number of groups of 14 (KO2). As seen by their work in Fig. 3, Blanca and Stuart both used multiplication but Blanca approached it through a partially-linear, tabular representation, while Stuart used single separated multiplications to narrow down the correct number of groups. Leon and Chioma's work are examples of the different kinds of division algorithms students enacted. Student strategies involving finding the difference (KO1) and the number of groups (KO2) will be discussed in more detail in their own sections below.

4.1.2. Partial evidence of sensemaking

One-hundred-fifteen students demonstrated partial evidence of sensemaking about the Fair Task through their attention to mathematical work for KO1, KO2, or both KO1 and KO2, but did not show evidence for KO3. Fig. 4 indicates that 115 out of the 280 students (41 %) provided evidence that they made sense of either the difference, the number of groups of 14, or both. However, these students did not demonstrate evidence of their understanding of the relationship between the number of groups of 14 and the difference of 1302 and 630, which represented the total value of the child tickets that were sold. These students' inability to make sense of KO3 means that, on the whole, they did not enact mathematical strategies that led to a solution for the Fair Task, but many of them did compute their partial solution correctly. The students in this sensemaking level exhibited eight mathematically different strategies (see Fig. 4). These strategies will be discussed in more detail in Section 4.2.

Examples of student work demonstrating evidence for partial understanding can be seen in Fig. 5. Saskia's work shows a student who found the necessary difference (KO1), but after that was not able to see the necessity of finding groups of 14 (KO2) nor the relationship between the difference and the number of groups of 14 (KO3). In Jeremy's work, we saw a student who wanted to find a number of groups of 14 (KO2), yet did not make sense that there is a portion of the sales that were for the children and therefore did not find the difference (KO1) and did not connect the difference to the number of groups (KO3). In Topanga's work, we see a rare example of a student who showed evidence of having made sense of both the difference (KO1) and the number of groups of 14 (KO2) but selected the difference as the answer to the question of how many child tickets. Taken collectively, these examples help to illustrate how students made partial sense of the problem; however, were not able to arrive at the correct mathematical result. Within the partial level of sensemaking, there were 22 students out of 115 (19.1 %) who, like Saskia, displayed a sense of understanding KO1, but then moved on to unrealistic solutions involving, addition, subtraction, multiplication, or division. These strategies do not connect in any way to the situation found in the problem and appear as random attempts to find something that "may" work out serendipitously.

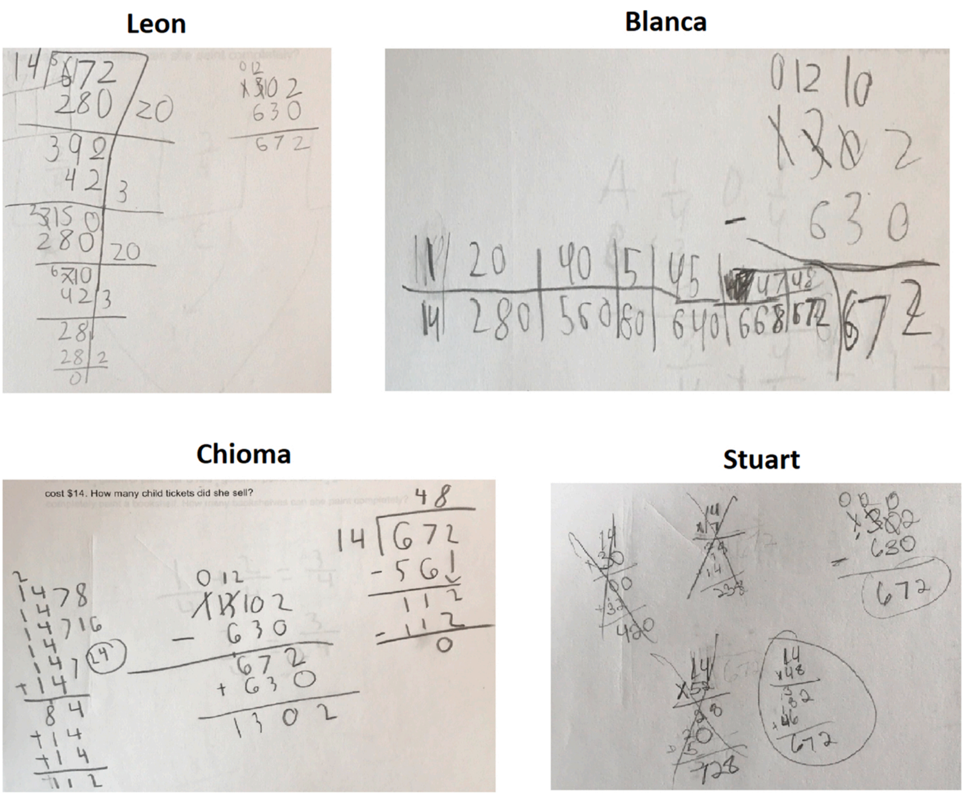

Fig. 3. Student work samples of different strategies for robust sensemaking of the Fair Task.

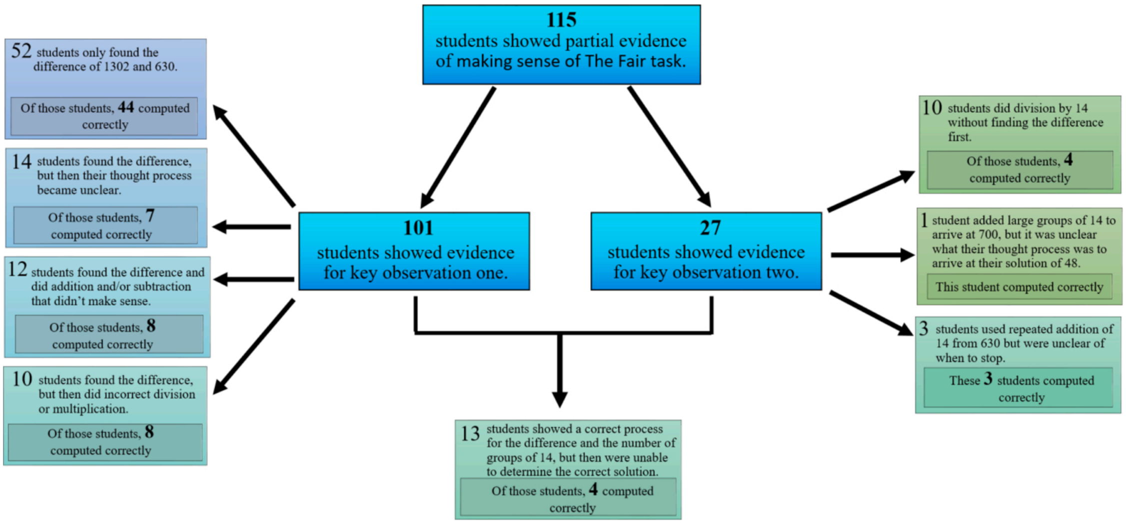

Fig. 4. Mapping of strategies for students who demonstrated partial evidence of sensemaking in their solving of the Fair Task. In this figure we show that there were 115 total students who showed evidence of partial sensemaking. Although most of the students in this category demonstrated evidence in either KO1 **or** KO2, there were 13 students who showed evidence in KO1 **and** KO2. Therefore, those 13 are accounted for in both the 101 students for KO1 and the 27 students for KO2.

4.1.3. *No evidence of sensemaking*

Students who displayed no evidence for any of the three key observations were classified in this level of sensemaking. As seen in Fig. 6, 86 out of the 280 students (31 %) provided no evidence that they had made sense of any of the three key observations. Students in this level of sensemaking either enacted strategies that would not lead to a correct solution of the Fair Task or gave no evidence of how they arrived at their solution. The majority of students in this level of sensemaking, 72 out of 86 (83.7 %), used numbers and operations within the problem in ways that did not align with the problem's context. For example, 24 of these students subtracted various numbers given in the problem (e.g., $1302-14 = 1288$ or $630-14 = 601$).

In Fig. 7, examples of these unrealistic computations are shown. Scott, Mackenzie, and Noah, show computations that are disconnected from the situation of the Fair Task as they tried to find the number of groups of 630 in 1302 (Scott), add the ticket prices together and then add that to the total money collected (Mackenzie), or multiply the ticket prices by the total amount of the adult ticket sales (Noah). Others found the number of groups of 18 within 1302 or 603. When students of this level of sensemaking demonstrated a mathematical computation, it revealed that they greatly struggled to connect it back to the situation within the problem (i.e., de-contextualize), limiting their ability to enact a viable strategy for any of the KOs.

4.2. *RQ2: students' strategy use on the fair task*

4.2.1. *Strategic choices for finding the difference*

As students made sense of KO1 involving the difference between 1302 and 630, they had representational and procedural choices to make. Three strategies were identified in students' solutions: (1) standard USA subtraction algorithm, which involved a symbolic representation to perform vertical subtraction by the take away model; (2) determining the difference by adding up, which involved a symbolic representation of adding up from 630 to arrive at 1302; and (3) number line, which involved determining the difference via pictorial representation of adding up from 630 to arrive at 1302 using a number line. The standard USA subtraction algorithm was the most prevalent strategy among students as it was used by 159 of 174 students who made sense of KO1. Adding up from 630, a much less prevalent strategy than the standard subtraction algorithm, was used by 14 of 174 students who made sense of KO1. Number line was only used by one student. Although the majority of students who made sense of KO1 used the standard algorithm, a small amount of procedural flexibility among students was observed through the 15 students who found the difference by alternative means. These students recognized that they could arrive at the difference by determining how much they needed to add to 630 to arrive at 1302.

4.2.2. *Strategic choices for finding the number of groups*

Students used a variety of methods to find the number of groups of 14 to represent the number of child tickets sold (KO2). Those procedures included: (1) repeated subtraction and addition; (2) multiplication (e.g., box method); (3) partial quotients; and (4) standard USA division algorithm. At times, the use of the standard division algorithm involved supporting computations of repeated subtraction, repeated addition, and/or multiplication. There were 106 of 280 (38 %) students who provided evidence of sensemaking about KO2. The number of students using these computation strategies was more uniformly distributed than those involving the difference (KO1). However, the standard USA division algorithm was still the most prevalent with 48 out of 106 students choosing to use it to find the number of groups of 14. The other three strategy types had less than half as many students choose to use them.

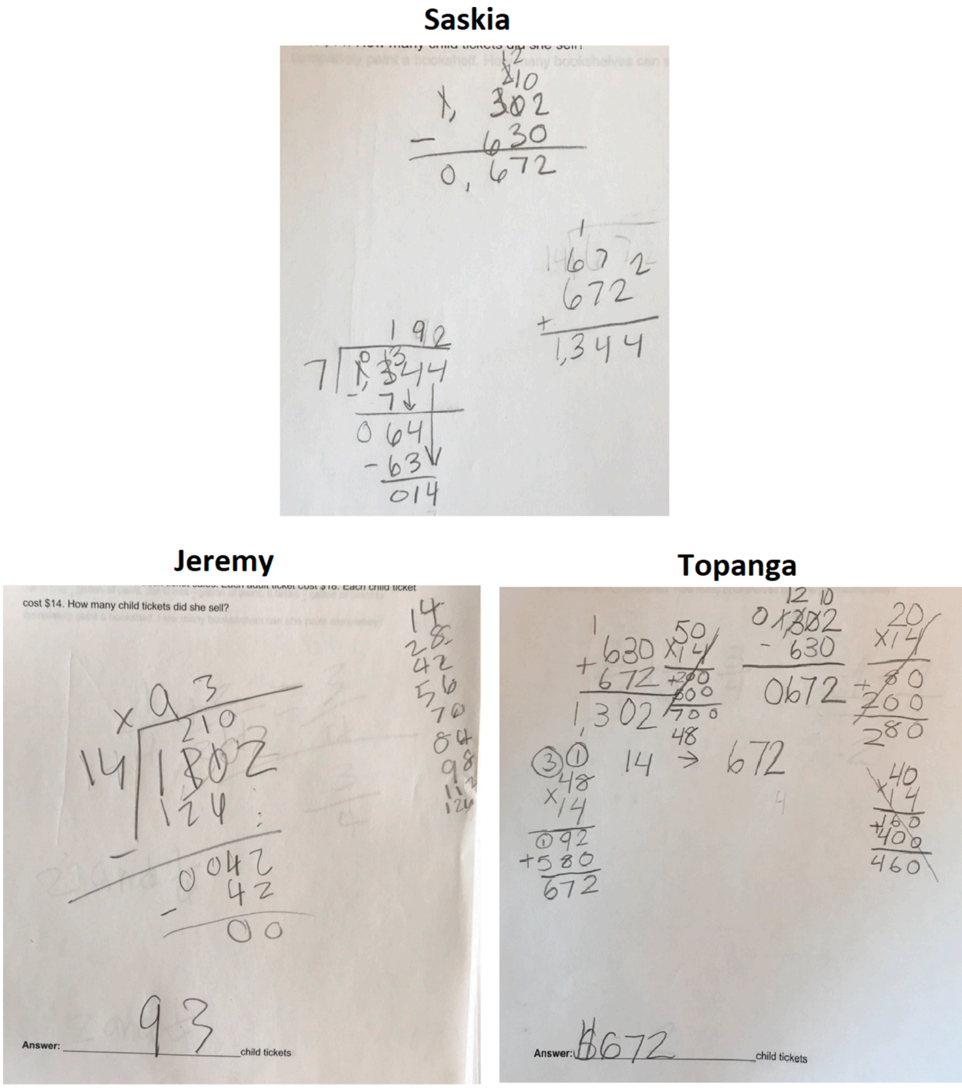

Fig. 5. Student work samples of different strategies for partical sensemaking of the Fair Task.

Repeated subtraction and addition, multiplication, and partial quotients, had 17, 18, and 23 students out of 106, respectively. Differences in strategies were found depending on whether students had a robust ($n = 79$) or partial ($n = 27$) level of sensemaking as can be seen in Figs. 1 and 3.

4.2.3. *Strategic choices for connecting the difference and number of groups*

All students who demonstrated sensemaking of KO3 had also given evidence for sensemaking of KO1 and KO2 and showed robust sensemaking of the Fair Task. Thus, student strategies for demonstrating how the difference and number of groups of 14 connected within the problem situation depended upon their strategies for both KO1 and KO2. Fig. 3 is a representative sample of the student strategies resulting in evidence of this connection. Observing the work of Leon, Blanca, Chioma, and Stuart shows how the vast majority (70 out of 79) of the robust sensemaking students used the standard USA subtraction algorithm to find the difference of 672 and then began to use many different mathematical strategies for finding the number of groups of 14 within 672. For example, Leon used partial quotients, Blanca used a multiplicative table of groups of 14, Chioma used the standard USA division algorithm, and Stuart used guess and check multiplication to find the number of groups of 14 within 672. In summary, student strategies for KO3 involved finding the difference in the exact same way, but from there diverged into a variety of strategies to find the number of groups of 14 in 672.

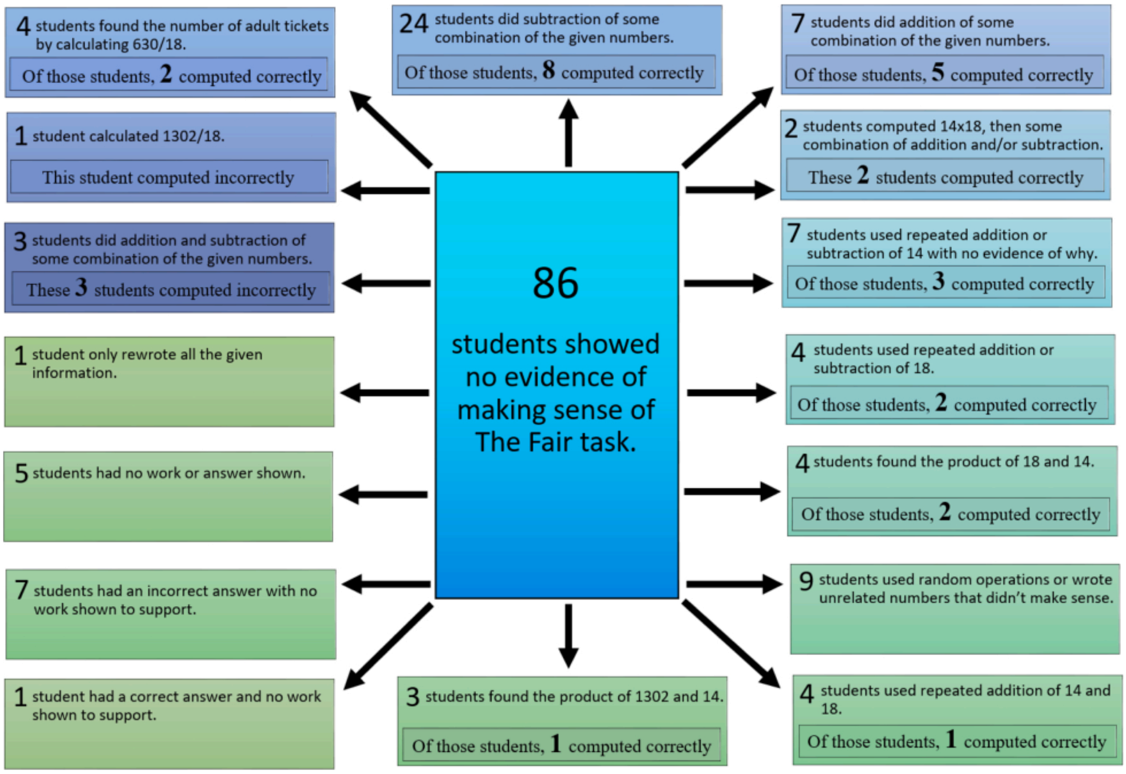

Fig. 6. Mapping of strategies for students who demonstrated no evidence of sensemaking in their solving of the Fair Task. Although none of the students in this level of sensemaking exhibited a related successful strategy, they did show unrelated computations. In this figure we share whether or not students successfully computed their chosen calculations that were unrelated to sensemaking of the problem.

4.3. RQ3: strategical success on the fair task

4.3.1. Student strategical success for finding the difference

In the Fair Task, students encountered a difference between a four-digit number and three-digit number that required at least one regrouping if done by the standard USA subtraction algorithm. Most students tended to be more successful in finding the difference when applying the standard subtraction algorithm as part of their strategy. On the other hand, finding the difference proved difficult for 20 students who recognized KO1 was necessary but made errors in their procedure. Of the 20 observed errors, 16 occurred with the standard subtraction algorithm and four occurred with adding up. Four types of errors were observed: (1) Copying, (2) Regrouping, (3) Arithmetic, and (4) Unknown. Four students made a copying error, indicating they wrote down one of the key numbers incorrectly. For example, one student wrote 630 as 603 in their subtraction. Eight students made a regrouping error while using the standard subtraction algorithm. Four students made an arithmetic error. For example, a student might have shown that $10-3=9$ in the process of finding the difference with the standard algorithm. Four students made an unknown error, meaning they did not show enough work to denote the type of error. For example, one student wrote $1302-630 = 668$. This incorrect result is probably a combination of regrouping errors, however, not enough evidence was given to confirm such a categorization.

4.3.2. Student strategical success for finding the number of groups

Table 2 shows the chosen strategies used by participants who showed robust evidence of sensemaking including the number of participants who found the correct number of groups of 14 there are in 672. The evidence indicated that students who demonstrated robust sensemaking were most successful with finding the number of groups when using multiplication. They were about equally likely to be successful when using the standard algorithm or partial quotients to divide and were least successful when using repeated addition and subtraction. Nine students used repeated subtraction or addition to try to find the number of groups. Eight of the nine students were unsuccessful as they had difficulty with organization and would denote that they were frustrated by the number of calculations they needed to do. Some students used shortcuts, such as adding three groups of 14 instead of just one. Though the students who used addition and subtraction perhaps chose to use an operation that they were more comfortable with, these operations were also less efficient. The students' lack of flexibility to switch to a more efficient strategy using multiplication or division inhibited their success. On the other hand, the 10 students who chose to use multiplication with an educated guess and check approach all succeeded in finding the correct answer. From our observations of students' work, this approach required far fewer calculations than addition and subtraction, and showed students' flexibility to use multiplication facts familiar to them.

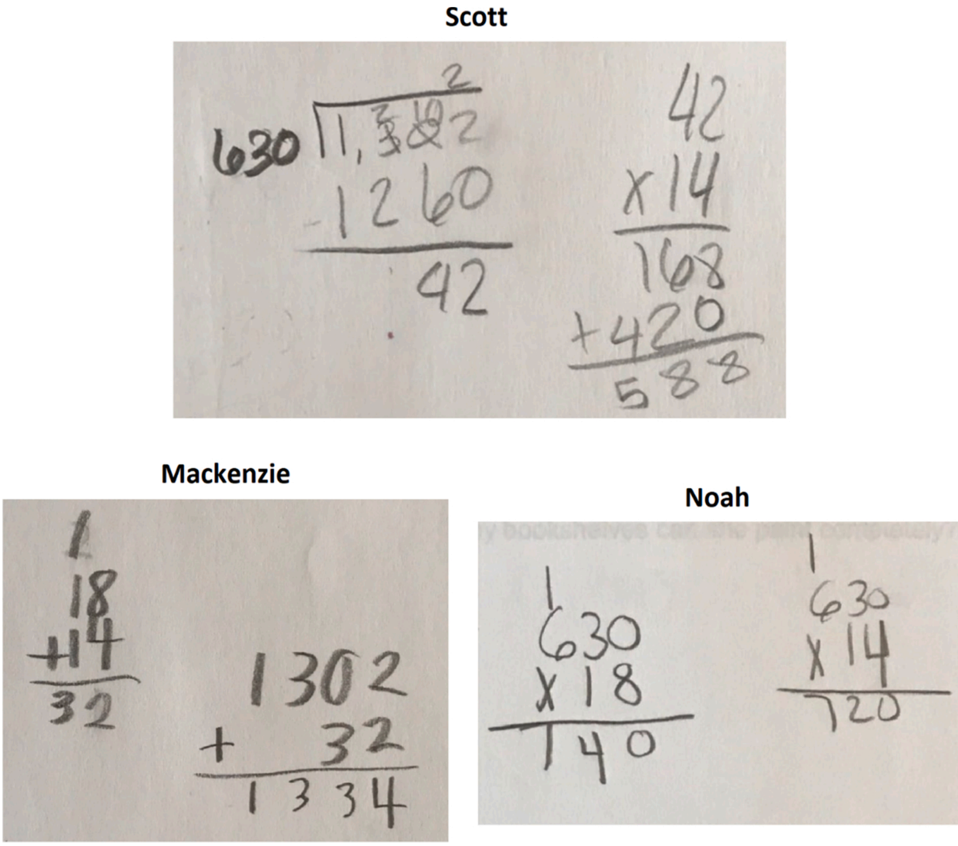

Fig. 7. Student work samples of different strategies for no sensemaking of the Fair Task.

Table 2
Student Strategies attending to Key Observation 2 with rates of computational success.

Strategy	Level of Sensemaking	Amount of Participants	Number of Participants with Correct Computation	Percentage Correct
Repeated addition and subtraction	Robust	9	2	22.22%
	Partial	8	4	50%
Multiplication	Robust	11	11	100%
	Partial	7	3	42.86%
Standard Division Algorithm	Robust	39	27	69.23%
	Partial	9	4	44.44%
Partial quotients	Robust	20	12	60%
	Partial	3	1	33.33%

Students who showed robust evidence of sensemaking but made computation errors did so in a variety of ways. One error type was organizational. Students did not accurately keep track of the number of groups of 14 when using repeated addition and subtraction. There were fifteen students who made errors using the standard division algorithm while finding the number of groups of 14. Five of those errors revealed that students were unsure of how to use the algorithm, which we termed algorithmic errors. Other categorized errors while executing the standard division algorithm included seven students who made arithmetic errors, two students who copied their answer incorrectly, and one student who showed understanding of how to use the algorithm but did not align numbers appropriately. Students who used the partial quotients strategy made errors involving the remainder, which was a result of their miscalculation of the number of groups of 14. Two students added the quotient and remainder, whereas three students listed a remainder as part of their answer showing a lack of attention to making sense of the remainder in context of the problem.

Table 2 also shows the differing success rates among students' strategies for KO2. Those students exhibiting partial evidence had not made full sense of the Fair Task like their peers, and this inhibited their success in finding the correct solution. Though these students gave evidence that they knew they needed to find a number of groups of 14, they were unable to connect the number of groups to the difference of 672 (KO3). Therefore, the numbers of correct computations listed in Table 2 for the partial sensemaking level do not reflect the number of students who found the correct answer of 48. Rather, the percentage of correct computations means for their chosen partial strategy, their computations for finding a number of groups of 14 were correct. If these students had been able to make sense of the whole problem, their computational prowess makes it more probable that they would have derived a correct solution to the overall problem. The largest inhibiting factor for these students was their inability to make sense of KO3. Looking inclusively at the computational efforts of all the students who demonstrated sense making for finding the number of groups of 14, we can derive from Table 2 that 64 out of 106 of them correctly computed. That is to say, about 60 % of the students who made sense of KO2 went on to demonstrate correct calculations involving finding a number of groups of 14.

4.4. Summary of findings

The number of different strategic choices made by students demonstrated the open nature of the Fair Task. Various strategies used illuminated differences in students' sensemaking and response to a multi-step word problem. Out of the 174 students who gave evidence for making sense of KO1, 159 of them used the standard algorithm for subtraction, including all of the students who showed robust evidence of sensemaking. In contrast, KO2 opens up the possibility for division to be used as the students needed to find the number of groups of 14 that go into 672. However, the students in this study used all four operations to make sense of and solve KO2. Students' strategies when sensemaking about KO2 showed that many understood they were able to use properties of operations and the relationships between the operations in their quest to find the number of groups. Yet, there were differences in the rates of success among the strategies as only two-thirds of the students using algorithmic processes for division proceeded to get the correct answer while 100 % of the students using multiplication methods to arrive at the number of groups of 14 arrived at the correct answer. Lastly, only 28 % of all students were able to demonstrate sensemaking about the connection between KO1 and KO2 and this greatly restricted the number of students who could successfully solve the problem.

5. Discussion and implications

Expectations for students doing two-step problem solving begins in second and third grade (CCSSI, 2010, p. 19, 23). By fourth-grade, the expectations rise to multistep word problem solving as indicated by 4.OA.3, "Solve multistep word problems posed with whole numbers and having whole-number answers using the four operations, including problems in which remainders must be interpreted (CCSSI, 2010, p. 29)." The Fair Task appears to offer a substantial challenge for student sensemaking, as it necessitates finding a difference (KO1), finding a number of groups of 14 (KO2), and understanding the relationship between the two (KO3). Past research has demonstrated that problem solving is difficult for elementary, middle grades, and secondary students (e.g., Bostic et al., 2017; Folger & Bostic, 2015; Verschaffel et al., 1999). While only 18 % of participants solving the problem might seem concerning, it is reasonable within the context of word problem solving. Word problem solving is much more difficult than solving mathematical exercises (Kilpatrick et al., 2001).

Solving a multi-step word problem involving well known operations may be more difficult for students due to the problem complexity. Unpacking this further, making sense of word problems involving addition and subtraction *should be* extremely familiar to fourth-grade students as it is an expectation of each prior grade, including Kindergarten (CCSSI, 2010, pp. 11, 15, 19, 23). The results of this study showed that when students are given a multi-step word problem, 106 of the 280 students (37.9 %) were unable to make sense of a context involving a difference (KO1), which could have been found by either addition or subtraction. Of the 174 students who displayed evidence of sensemaking for KO1, 20 made minor errors which kept them from finding the correct result. This means that there were more than five times as many students who struggled with the sensemaking as there were students who made computational errors. Hence, struggles with sensemaking account for the largest reason why students did not complete KO1 successfully. The evidence here suggests that although students have had nearly five years of formal education on addition and subtraction, when encountering those operations in the complexity of a multi-step word problem, many of them struggle to connect the context of the problem to those operations.

In their consideration of the two subtraction models, take away and determining the difference, Selter et al. (2012) found that "Using the determining the difference model and the inverse relation between addition and subtraction seems to be efficient; however, it is rarely being used" (p. 404). They further articulated that this imbalance exists in spite of the empirical research evidence showing determining the difference's application to mental strategies, application to all algorithms, and more robust extension to future mathematical ideas. The research here continues the line of evidence that take away models were most prevalent among students as 159 of the 174 students who made sense of the difference in the Fair Task used the standard USA subtraction algorithm (take away model) as part of their strategy. Although it is beyond the scope of this study to make claims about students' instructional environments, we note that mathematics educators guided by the CCSSM's movement toward standard algorithm mastery in fourth grade should be aware of the possible imbalance between the two subtraction models being enacted through instruction. Further research on the connection between standards and instruction could illuminate the status of student engagement in the two models of subtraction.

Making sense of word problems involving multiplication and division is expected in grades 3 and 4 (CCSSI, 2010, pp. 22, 29). The finding here showed that when students are engaging in a multi-step word problem containing a key observation involving the number of groups (KO2), 59 % of students were unsuccessful in their sensemaking. Furthermore, drawing from Table 2, we noted that among those students who did make sense of KO2, only 60 % of them went on to demonstrate correct computations. This adds to the findings of other studies which have shown that computing the number of groups is difficult for upper elementary students (Fagginger Auer et al., 2016; Hickendorff et al., 2009).

Accounting for those who were unable to make sense of the connection between the difference and the number of groups (KO3) increases the percentage of students unable to demonstrate a complete sensemaking of the Fair Task to 71.8 %. The implication of this finding adds to the current literature because it demonstrates a large difficulty for students during multistep problem solving. That being, when students engage in a multi-step word problem, they struggle greatly in determining the connection between two different mathematical operations within the context of the problem. While our findings contribute additional support for student difficulties in sensemaking as shown by other studies (Matney et al., 2013; Palm 2008, Pape, 2004; Yee & Bostic, 2014), it provides additional evidence of the difficulty that students have in making sense of a difference (KO1), a number of groups (KO2), and the connection between the two (KO3) when encountering these elements in a multi-step word problem. As students persevere through multi-step word problem solving, mathematics educators should seek to devise instruction that supports students in not only making sense of two or more operations like subtraction and division, but also how the results of those operations fit together to generate a reasonable outcome for the problem (i.e., contextualization).

5.1. *Familiar and efficient strategies*

In contrast to the expectation of student fluency for the standard subtraction algorithm, students in the fourth-grade are still learning about division and are not yet expected to fluently apply the standard division algorithm. In the Fair Task, students needed to find the number of groups of 14 found in 672 (KO2), and this opened up a possibility for division. The fourth-grade expectations for division include being able to find whole-number quotients "using strategies based on place value, the properties of operations, and/or the relationship between multiplication and division" (CCSSI, 2010, p. 30). Students in this study used all four operations to come to a solution (see Table 2). Students' strategies when making sense of KO2 showed that many understood they were able to use properties of operations and the relationships between the operations to find the quotient. Among those who had robust sense making, the highest rates of success came from those who devised various multiplicative strategies, such as Blanca and Stuart (Fig. 3). Other students who chose to add, subtract, or use a division algorithm succeeded at lower rates. This seems to support Torbeyns et al.'s (2018) claim that the most successful students adapt their strategy choices to the problem at hand. Similarly, these findings extend from Yee & Bostic (2014) who found that sixth- and tenth-grade students who used multiple strategies while solving a word problem, inclusive of symbolic and/or nonsymbolic approaches, had greater problem-solving performance than peers who solely used a symbolic approach like an algorithm. In their prior research, sensemaking artfully informed students' strategy use and in turn improved how well they were able to solve word problems. The implication from the present study is that students who chose to use a familiar, well understood, and efficient operational strategy, had higher rates of success than students who used less efficient strategies *and* students who used digit-based methods, such as division algorithms.

5.2. *Algorithmic outcomes for division*

The majority of students who gave evidence for robust sensemaking approached KO2 using an algorithm to find the number of groups; 39 used the standard division algorithm and 20 used partial quotients. As indicated in Table 2, students using these two algorithms showed about the same levels of success in arriving at the correct calculation for the number of groups. Although we found that a similar number of students were successful with both the standard division algorithm and partial quotients, students using the partial quotients algorithm demonstrated flexibility in choosing values for finding the quotient. These observations seem to support Kilpatrick et al.'s (2001) claim that the partial quotients method of division provides accessibility for students by allowing them the opportunity to choose their own values needed to find the number of groups.

5.3. *Tendency for unrealistic solutions*

Findings from this study align with others, suggesting that one of the reasons students give unrealistic answers is that they enact "mindless calculations and do not include considerations of the real life aspects of the situations described in the tasks" (Palm, 2008, p. 38). We noticed that when students were unable to make sense of the Fair Task, they most often demonstrated seemingly random calculations to relate the quantities without a purposeful strategy or knowing the meaning of the operation. In doing so, these students seem to leave their chances of obtaining the correct solution to serendipity. This finding connects with Verschaffel and colleague's (2000) finding that when solving arithmetic word problems in school settings, students tended to use one or more arithmetic operations involving the quantities in the problem without consideration of the context. Our study extends their finding by observing the levels of sensemaking and the rate of this phenomena within each level. That being, 19.1 % of students in the partial sensemaking level and 83.7 % of students in the no sensemaking level enacted unrealistic solutions. As such, it remains difficult to discern if much has changed with regard to student sensemaking about word problems. The primary difficulty found here was students' ability to connect the situation to an appropriate mathematical procedure, which is the decontextualization (CCSSI, 2010) aspect of sensemaking.

6. Future research and limitations

The study here looked at 280 fourth-grade students' sensemaking and strategies for a multistep word problem addressing the Operations and Algebraic domain of the CCSSM (CCSSI, 2010). As such, the findings shared here about student sensemaking should be tempered toward this domain. The focus on fourth-grade students describes a picture of what happens in the continuum of their education and is not meant to describe the sensemaking at all ages or in all mathematical aspects of elementary children. Similarly, the Fair Task involves opportunities for students to enact the four operations most focused on during the elementary years. However, as seen by the KOs, the task is particularly germane to understanding how students make sense of a difference, a number of groups, and connect those two within the context of the problem. Therefore, further research is needed on students' sensemaking in other content domains, students at other grade levels, and other multistep arithmetic word problem types involving different combinations of operations.

7. Conclusion

We have provided analysis and mappings of students' sensemaking and related strategies for solving a multistep word problem. Sensemaking remains a primary struggle for many students solving word problems. When engaging with the complexities of a multistep word problem, students were most successful making sense of the difference (KO1; 62.1 %), but less successful making sense of the number of groups (KO2; 37.9 %) and the least successful connecting the difference and the number of groups (KO3; 28.2 %). Student sensemaking remains an area of difficulty in need of attention. The various strategies used by fourth-grade students illuminated differences in how they made sense of and responded to a multistep word problem. We observed that fourth-grade students are better able to make sense of the contextual parts of the word problem for operations, like addition and subtraction (KO1), that they have studied for a long time. Furthermore, among those who robustly made sense of context, they overwhelmingly chose the standard subtraction algorithm, though a few students still struggled to use it. Still more, it is problematic that more than a third of the students showed no signs of sensemaking of finding a difference (KO1) at all. Fourth-grade students were less able to make sense of a context asking them to find a number of groups (KO2) but, as a set of participants, they showed much more flexibility in their strategies to figure out the number of groups than they did the difference. For those who exhibited robust sensemaking, the strategy of employing multiplication was found to be more successful than division algorithms. Overall, sensemaking about the context appears to be the area of greatest struggle. For students to be successful sense makers of arithmetic word problems, they must have conceptual understanding about the operations, be able to connect those understandings to the context, and then enact their mathematical operations correctly according to their chosen strategic purpose. The mathematics education community has been considering these ideas for many decades (NCTM; 1980, 1989, 2000, 2007, 2009, 2014, 2020) and the evidence here shows that much work remains for us as we seek to help students become better mathematical problem solvers. The findings here support Resnick' s (1988) assertion that becoming a good problem solver is as much a matter of sensemaking as it is acquiring any particular set of skills or strategies. More research is needed on how to help students grow in all of these areas.

Data availability

Data will be made available on request.

Funding

This work was supported by the National Science Foundation [NSF #1720646; #1720661; #2100988; #2101026]. Any opinions, findings, conclusions, or recommendations expressed by the authors do not necessarily reflect the views of the National Science Foundation.

Role of the funding source

The National Science Foundation (NSF) had no involvement in the study design, the collection, analysis, and interpretation of data. Similarly, NSF had no involvement in the writing of this manuscript and the decision to submit for publication.

Submission declaration

The aforementioned authors declare that this work has not been published previously, is not under consideration for publication elsewhere, is approved by all authors to be submitted to JMB according to the agreed author order above, and will not be published elsewhere.

CRediT authorship contribution statement

Gabriel Matney: Conceptualization, Methodology, Supervision, Formal analysis, Writing – original draft. **Jonathan D. Bostic:** Writing – original draft, Formal analysis, Funding acquisition. **Miranda Fox:** Writing – original draft, Visualization, Data curation, Formal analysis. **Tiara Hicks:** Writing – original draft, Visualization, Data curation, Formal analysis. **Toni May:** Writing – review & editing, Funding acquisition. **Greg Stone:** Writing – review & editing.

Declaration of Competing Interest

Ideas in this manuscript stem from grant-funded research by the National Science Foundation (NSF #1720646; #1720661, #2100988, #2101026). Any opinions, findings, conclusions, or recommendations expressed by the authors do not necessarily reflect the views of the NSF.

Acknowledgements

We appreciate the review and feedback from Dr. Karl Kosko during the final preparations of this manuscript.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Author.

Anghileri, J., Beishuizen, M., & van Putten, K. (2002). From informal Strategies to structured procedures: Mind the gap! *Educational Studies in Mathematics, 49*(2), 149–170.

Baroody, A. J., & Dowker, A. (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise*. Mahwah, NJ: Erlbaum.

Becker, J. P., & Shimada, S. (1998). *The open-ended approach*. National Council of Teachers of Mathematics.

Bostic, J., & Sondergeld, T. (2015). Measuring sixth-grade students' problem solving: Validating an instrument addressing the mathematics Common Core. *School Science and Mathematics Journal, 115*, 281–291.

Bostic, J., Matney, G., Sondergeld, T., & Stone, G. (2018). Content validity evidence for new problem-solving measures (PSM3, PSM4, and PSM5). November. In T. Hodges, G. Roy, & A. Tyminski (Eds.), *Proceedings for the 40h Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (p. 1641).

Bostic, J., Matney, G., Sondergeld, T., & Stone, G. (2019). Developing a problem-solving measure for grade 4. July. In , *Proceedings of the 43rd Meeting of the International Group for the Psychology of Mathematics Education, 4* pp. 4–16).

Bostic, J., Matney, G., Sondergeld, T., & Stone, G. (2020). Measuring what we intend: A validation argument for the grade 5 problem-solving measure (PSM5). Validation: A burgeoning methodology for mathematics education scholarship. March. In J. Cribbs, & H. Marchionda (Eds.), *Proceedings of the 47th Annual Meeting of the Research Council on Mathematics Learning* (pp. 59–66).

Bostic, J., Pape, S., & Jacobbe, T. (2016). Encouraging sixth-grade students' problem-solving performance by teaching through problem solving. *Investigations in Mathematics Learning, 8*(3), 30–58.

Bostic, J., Sondergeld, T., Folger, T., & Kruse, L. (2017). PSM7 and PSM8: Validating two problem-solving measures. *Journal of Applied Measurement, 18*(2), 151–162.

Bostic, J., Sondergeld, T., Matney, G., Stone, G., & Hicks, T. (2021). Gathering response process data for a problem-solving measure through whole-class think alouds. *Applied Measurement in Education, 34*(1), 46–60.

Cai, J., & Lester, F. K. (2005). Solution representations and pedagogical representations in Chinese and U.S. classrooms. *Journal of Mathematical Behavior, 24*, 221–237.

Carmichael, S. B., Martino, G., Porter-Magee, K., & Wilson, W. S. (2010). *The state of the state standards and the Common Core in 2010*. Retrieved from: Thomas B. Fordham Institute https://fordhaminstitute.org/national/research/state-state-standards-and-common-core-2010.

Caviola, S., Mammarella, I. C., Pastore, M., & LeFevre, J.-A. (2018). Children's strategy choices on complex subtraction problems: Individual differences and developmental changes. *Frontiers in Psychology, 9*(1209), 1–16. https://doi.org/10.3389/fpsyg.2018.01209

Cifarelli, V. V., & Cai, J. (2005). The evolution of mathematical explorations in open-ended problem-solving situations. *The Journal of Mathematical Behavior, 24*, 302–324.

Common Core State Standards Initiative. (2010). *Common core standards for mathematics*. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.

Creswell, J. (2011). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.

Fagginger Auer, M. F., Hickendorff, M., van Putten, C. M., Béguin, A. A., & Heiser, W. J. (2016). Multilevel latent class analysis for large-scale educational assessment data: Exploring the relation between the curriculum and students' mathematical strategies. *Applied Measurement in Education, 29*(2), 144–159. https://doi.org/10.1080/08957347.2016.1138959

Folger, T., & Bostic, J. (2015f). Supporting middle school students' problem solving with the PSM6. *School Science and Mathematics Journal, 115*, 1–3. Retrieved from: https://www.ssma.org/assets/docs/Oct2015folgerbostic.pdf.

Goldin, G. (2002). Representation in mathematical learning and problem solving. In L. D. English, & M. G. Bartolini Bussi (Eds.), *Handbook of international research in mathematics education* (2nd ed.). New York: Routledge. pp. xiii, 925 p.

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

Hatch, J. A. (2002). *Doing qualitative research in educational settings*. Albany, NY: State University of New York Press.

Hickendorff, M. (2013). The effects of presenting multidigit mathematics problems in a realistic context on sixth graders' problem solving. *Cognition and Instruction, 31*(3), 314–344.

Hickendorff, M., Heiser, W. J., van Putten, C. M., & Verhelst, N. D. (2009). Solution strategies and achievement in Dutch complex arithmetic: Latent variable modeling of change. *Psychometrika, 74*(2), 331–350.

Hickendorff, M., Torbeyns, J., & Verschaffel, L. (2018). Grade-related differences in strategy use in multidigit division in two instructional settings. *British Journal of Developmental Psychology, 36*, 169–187.

Hickendorff, M., van Putten, C. M., Verhelst, N. D., & Heiser, W. J. (2010). Individual differences in strategy use on division problems: Mental versus written computation. *Journal of Educational Psychology, 102*(2), 438–452.

Imbo, I., Vandierendonck, A., & Vergauwe, E. (2007). The role of working memory in carrying and borrowing. *Psychological Research, 71*, 467–483. https://doi.org/10.1007/s00426-006-0044-8

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Koestler, C., Felton, M. D., Bieda, K. N., & Otten, S. (2013). *Connecting the NCTM process standards and the CCSSM practices*. Reston, VA: National Council of Teachers of Mathematics.

Levingston, H. B., Neef, N. A., & Cihon, T. M. (2009). The effects of teaching precurrent behaviors on children's solution of multiplication and division word problems. *Journal of Applied Behavior Analysis, 42*(2), 361–367. https://doi.org/10.1901/jaba.2009.42-361

Linacre, J. M. (2021). *A users guide to winsteps ministep rasch-model computer programs* [Manual 5.1.0]. Retrieved from https://www.winsteps.com/manuals.htm.

Matney, G., Jackson, J., & Bostic, J. (2013). Effects of minute contextual experience on realistic assessment of proportional reasoning. *Investigations in Mathematics Learning, 6*(1), 41–68.

Mayer, R., & Wittrock, M. (2006). Problem solving. In P. Alexander, & P. Winne (Eds.), *Handbook of educational psychology* (pp. 287–303). Mahwah, NJ: Erlbaum.

Mueller, M., Yankelewitz, D., & Maher, C. (2011). Sense making as motivation in doing mathematics: Results from two studies. *The Mathematics Educator, 20*(2), 33–43.

National Council of Teachers of Mathematics. (1980). *An agenda for action*. 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. (2007). Mathematics teaching today: Improving practice. In T. Martin (Ed.), *Improving student learning* (2nd ed.). Reston, VA: Author.

National Council of Teachers of Mathematics. (2009). *Focus in high school mathematics: Reasoning and Sense making*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2014). *Principles to action*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2020). *Catalyzing change in early childhood and elementary mathematics: Initiating critical conversations*. Reston, VA: Author.

Nortvedt, G. A. (2010). Understanding and solving multistep arithmetic word problems. *Nordic Studies in Mathematics Education, 15*(3), 23–50.

Palm, T. (2006). Word problems as simulations of real-world situation: A proposed framework. *For the Learning of Mathematics, 26*, 42–47.

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics, 67*, 37–58.
Pape, S. (2004). Middle school children's problem-solving behavior: A cognitive analysis from a reading comprehension perspective. *Journal for Research in Mathematics Education, 35*(3), 187–219.
Quintero, A. H. (1983). Conceptual understanding in solving two-step word problems with a ratio. *Journal for Research in Mathematics Education, 14*(2), 102–112.
Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmarks Paedagoiske Institut.
Reed, S. K. (1999). *Word problems: Research and curriculum reform*. Mahwah, NJ: Erlbaum.
Resnick, L. (1988). Treating mathematics as an ill-structured discipline. In R. Charles, & E. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 32–60). Reston, VA: National Council of Teachers of Mathematics.
Sahin, N., Dixon, J. K., & Schoen, R. C. (2020). Investigating the association between students' strategy use and mathematics achievement. *School Science and Mathematics, 120*(6), 325–332. https://doi-org.ezproxy.bgsu.edu/10.1111/ssm.12424.
Schindler, M., & Bakker, A. (2020). Affective field during collaborative problem posing and problem solving: A case study. *Educational Studies in Mathematics, 105*, 303–324. https://doi.org/10.1007/s10649-020-09973-0
Schmidt, W. H., & Houang, R. T. (2012). Curricular coherence and the common core state standards for mathematics. *Educational Researcher, 41*(8), 294–308.
Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). New York: Macmillan.
Schoenfeld, A. H. (2011). *How we think: A theory of goal--oriented decision making and its educational applications*. New York, NY: Routledge.
Schulz, A., & Leuders, T. (2018). Learning trajectories towards strategy proficiency in multi-digit division – A latent transition analysis of strategy and error types. *Learning and Individual Differences, 66*, 54–69.
Selter, C. (2001). Addition and subtraction of three-digit numbers: German elementary children's success, methods and strategies. *Educational Studies in Mathematics, 47*(2), 145–173.
Selter, C., Prediger, S., Nührenbörger, M., & Hußmann, S. (2012). Taking away and determining the difference—A longitudinal perspective on two models of subtraction and the inverse relation to addition. *Educational Studies in Mathematics, 79*, 389–408. https://doi.org/10.1007/s10649-011-9305-6
Smith-Chant, B. L., & LeFevre, J.-A. (2003). Doing as they are told and telling it like it is: Self-reports in mental arithmetic. *Memory & Cognition, 31*, 516–528. https://doi.org/10.3758/BF03196093
Sowder, L. (1985). Cognitive psychology and mathematical problem solving: A discussion of Mayer's paper. In E. A. Silver (Ed.), *Teaching and learning mathematical problem-solving: Multiple research perspectives* (pp. 139–145). Lawrence Erlbaum.
Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2006). The development of children's adaptive expertise in the number domain 20 to 100. *Cognition & Instruction, 24*(4), 439–465. https://doi.org/10.1207/s1532690xci2404_2
Torbeyns, J., Peters, G., De Smedt, B., Ghesquiere, P., & Verschaffel, L. (2018). Subtraction by addition strategy use in children of varying mathematical achievement level: A choice/no-choice study. *Journal of Numerical Cognition, 4*(1), 215–234.
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., Greer, B., & DeCorte, E. (2000). *Making sense of word problems*. Lisse, Netherlands: Swets & Zeitlinger.
Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM Mathematics Education, 52*, 1–16. https://doi.org/10.1007/s11858-020-01130-4
Verschaffel, L., Greer, B., Van Dooren, W., & Mukhopadhyay, S. (2009). *Words and worlds: Modelling verbal descriptions of situations*. Rotterdam: Sense Publishers.
Verschaffel, L., Greer, B., & Decorte, E. (2007a). Whole number concepts and operation. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 557–628). Charlotte, NC: Information Age Publishing.
Verschaffel, L., Torbeyns, J., De Smedt, B., Luwel, K., & Van Dooren, W. (2007b). Strategy flexibility in children with low achievement in mathematics. *Educational and Child Psychology, 24*(2), 16–27.
Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*, 458–477.
Yee, S., & Bostic, J. (2014). Developing a contextualization of students' mathematical problem solving. *Journal for Mathematical Behavior, 36*, 1–19.
Yoshida, H., Verschaffel, L., & De Corte, E. (1997). Realistic considerations in solving problematic word problems: Do Japanese and Belgian children have the same difficulties? *Learning and Instruction, 7*(4), 329–338.