Journal for Research in Mathematics Education
2018, Vol. 49, No. 5, 575–613

A Cross-Sectional Investigation of Students' Reasoning About Integer Addition and Subtraction: Ways of Reasoning, Problem Types, and Flexibility

Lisa L. Lamb
San Diego State University

Jessica Pierson Bishop
Texas State University

Randolph A. Philipp
San Diego State University

Ian Whitacre
Florida State University

Bonnie P. Schappelle
San Diego State University

In a cross-sectional study, 160 students in Grades 2, 4, 7, and 11 were interviewed about their reasoning when solving integer addition and subtraction open-number-sentence problems. We applied our previously developed framework for 5 Ways of Reasoning (WoRs) to our data set to describe patterns within and across participant groups. Our analysis of the WoRs also led to the identification of 3 problem types: change-positive, all-negatives, and counterintuitive. We found that problem type influenced student performance and tended to evoke a different way of reasoning. We showed that those with more experience with negative numbers use WoRs more flexibly than those with less experience and that flexibility is correlated with accuracy. We provide 3 types of resources for educators: (a) WoRs and problem-types frameworks, (b) characterization of flexibility with integer addition and subtraction, and (c) development of a trajectory of learning about integers.

Keywords: Addition; Children's thinking; Flexibility; Integers; Negative numbers; Open number sentences; Problem types; Subtraction

Consider second-grader Sam's and seventh-grader Ann's responses to the open number sentence $6 + \square = 4$.

This article is based on work supported by the National Science Foundation (NSF) under Grant No. DRL-0918780. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

Sam: Six plus blank equals 4. [Pause.] I can't do that one, either.
Int.: Why not?
Sam: Because that is 6, and how could it [the sum] equal 4 unless this [addition sign] was a minus? [Then] I could do it.
Int.: Can you explain why it doesn't work when it's plus?
Sam: Because, don't you see, it's 6. And if I plus more, how would I get 4? Even if I put zero right here [points to the blank], I couldn't get 4.
Ann: Six plus box equals 4. That would be negative 2.
Int.: How did you think about that?
Ann: Because . . . you can't have a positive number [points to the blank] to get a number that is less than the first number. So you would have to have a negative number right there [points to the blank].
Int.: So, I understand why it has to be negative. How did you know it was going to be a negative 2?
Ann: Because 6 is greater than 4 by 2. So, to actually get it down to 4 you would actually have to subtract 2. So it is like minus 2.

Sam's and Ann's responses showcase examples of two of five ways of reasoning documented in our cross-sectional investigation of students' understanding of integer addition and subtraction. For the same open number sentence, Sam's implicit generalization that when one adds two numbers, the sum should not be smaller than the addends, which is correct in the domain of whole numbers, is one that Ann appropriately leverages in her solution. As this brief example illustrates, the results of our investigation of students' thinking about integers provide insights into the emergent and powerful ways that students reason about integer addition and subtraction open number sentences both before and after school-based integers instruction. Moreover, the cross-sectional nature of our study enabled us to look across participant groups to analyze differences, similarities, and connections across grade levels.

Fluency with integer operations marks a transition from arithmetic to algebra and serves as a foundation for algebra because of the abstract nature of integers (Hefendehl-Hebeker, 1991; Linchevski & Williams, 1999; Peled & Carraher, 2008; Vlassis, 2002). For example, to navigate algebraic equations, students must perform algebraic procedures using additive inverses, which first come into play with the introduction of integers. Further, researchers have documented that students' difficulties with algebra often relate to their understanding of number, specifically integers (Moses, Kamii, Swap, & Howard, 1989; Vlassis, 2002). Fluency with integers can serve as an important milestone for students' future success in mathematics. Given the well-documented struggles that students have operating with negative numbers (e.g., Murray, 1985; Thomaidis & Tzanakis, 2007), understanding students' thinking about integer operations is particularly important to better support their learning. In reviewing the literature, we found no study in which elementary, middle, and high school students were sampled to document their reasoning about integer addition and subtraction. The findings from this study provide a cross-grades view of (a) students' ways of reasoning about integer addition and subtraction, (b) the relationships between problem types and ways of reasoning, and (c) the role that flexibility plays in solving addition and subtraction problems. The findings reported here advance the field's collective understanding of students' thinking about integer addition and subtraction and contribute to the efforts of the mathematics education research community to consider implications for instruction to enable students to successfully transition from arithmetic to algebra. In the following sections, we share the theoretical perspective and the literature base on which this work builds. We close this section by sharing a conceptual framework in which the relevant ideas and concepts that guide our study design, data collection, and analyses are integrated.

Literature Review

Students' Mathematical Reasoning: Our Theoretical Orientation

We approach our research from a children's mathematical thinking perspective (Carpenter, Fennema, Franke, Levi, & Empson, 2014; Case, 1996; Fuson, Smith, & Lo Cicero, 1997; Steffe, 1992, 1994, 2001; Steffe & Olive, 2010). Children's mathematical thinking is distinct from that of adults, and we take seriously the nature of that mathematics, whether correct or not from an expert perspective. We value seeing mathematics through children's eyes to better understand the sense they make. This perspective is grounded in Piaget's work and based on constructivist principles that children have existing knowledge and experiences they bring into the classroom and upon which they continue to build (e.g., Carpenter et al., 2014).

Generally speaking, the goal of this line of work within mathematics education has been to describe cognitive processes and structures that model or account for children's problem-solving activities in a given domain. Some scholars, like Case (1996), focused on more general cognitive structures, whereas others focused on domain-specific theories of student thinking (Carpenter & Moser, 1982, 1983; Steffe, 1992, 1994; Steffe & Olive, 2010). For example, Carpenter and Moser (1982, 1983) described increasingly abstract and efficient strategies for addition and subtraction problems—strategies that reflect the development of critical underlying cognitive processes. Within the constructivist paradigm, researchers have used a variety of constructs to study the development of student thinking, including schemes and operations (e.g., Steffe, 1992), cognitive processes reflected in student strategies (e.g., Carpenter & Moser, 1983), mental models and framework theory (e.g., Vosniadou & Brewer, 1992), and central conceptual structures (e.g., Case's, 1996, central numerical structure of the mental number line). Though these scholars differ in the specific goals of their work, the core constructs used, units of analysis, data sources, and analytic tools, this work has a common focus: to explain how students develop increasingly sophisticated mathematical understanding over time, whether through the accommodation or reorganization of schemes or mental models, increased flexibility in choosing different and more efficient strategies, the coordination of multiple schemes into larger cognitive structures, or operating with increasing levels of abstraction.

Although we share a broad constructivist orientation with others in this heritage, our approach to studying students' mathematical thinking in the realm of integers is most closely aligned with Carpenter and Moser's (1982, 1983) research on number concepts and the subsequent extensions of that work with the Cognitively Guided Instruction research program (Carpenter, Ansell, Franke, Fennema, & Weisbeck, 1993; Carpenter et al., 2014; Carpenter, Franke, & Levi, 2003; Empson & Levi, 2011). In work similar to theirs, we emphasize problem-solving processes and organizing student thinking into broader ways of reasoning and more-fine-grained strategies within these ways of reasoning. In our research, we focus on categorizing different approaches to integer addition and subtraction on the bases of patterns in students' solution strategies and the underlying views of number and operations at work in those strategies.

Additionally, we take this perspective because our ultimate goal in this research is to find ways to better support children's learning of mathematics. Instruction that is built on students' ideas benefits both teachers and students (Franke, Carpenter, Levi, & Fennema, 2001; Gearhart & Saxe, 2004; National Council of Teachers of Mathematics [NCTM], 2000; Sowder, 2007; Wilson & Berne, 1999). Moreover, instruction that is focused on responding to and developing students' mathematical thinking has been shown to support rich instructional environments (Gearhart & Saxe, 2004; NCTM, 2000; Sengupta-Irving & Enyedy, 2015; Sowder, 2007; Wilson & Berne, 1999) and contribute to gains in student achievement (Bobis et al., 2005; Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Fennema et al., 1996; Jacobs, Franke, Carpenter, Levi, & Battey, 2007; Villaseñor & Kepner, 1993). However, one must investigate, understand, and describe students' mathematical ideas before those ideas can be used meaningfully in instruction—thus, the need for our study.

Because we are interested in understanding students' conceptions related to integer addition and subtraction, we first review existing research about students' reasoning about integers and describe how we build on and extend this work in our study.[1] We also summarize research on problem types and flexibility because of our increasing understanding of the role that the two constructs play in understanding students' ways of reasoning, in general (Kilpatrick, Swafford, & Findell, 2001), and in understanding students' ways of reasoning about integers, specifically.

Students' Reasoning on Integer Addition and Subtraction Problems

Research involving students postinstruction. Researchers have investigated students' reasoning and performance on integer addition and subtraction problems

[1] We identified four additional areas of integers research that are beyond the scope of this paper: the historical development of integers, integer instruction, students' understanding of integer comparisons, and students' understanding of algebraic expressions and equations. For a comprehensive discussion of the historical development of integers, see Bishop, Lamb, Philipp, Whitacre, Schappelle, and Lewis (2014); Hefendehl-Hebeker (1991); and Henley (1999). For thorough overviews of the research related to integer instruction, see the literature reviews in Bofferding (2014) and Stephan and Akyuz (2012).

both after school-based instruction (e.g., Bofferding & Richardson, 2013; Chiu, 2001; Lamb et al., 2017; Murray, 1985; Vlassis, 2002, 2008) and before (e.g., Bishop, Lamb, Philipp, Whitacre, Schappelle, & Lewis, 2014; Bofferding, 2010; Murray, 1985; Peled, Mukhopadhyay, & Resnick, 1989). Studies of student reasoning conducted after the participants had experienced school-based integer instruction can be grouped into two categories on the basis of the age of the participants: college students (Bofferding & Richardson, 2013; Chiu, 2001) and middle or high school students who had experienced instruction with integers within 2 years (Chiu, 2001; Lamb et al., 2017; Murray, 1985). In the first category, college students, regardless of major, demonstrated high rates of success on both integer addition and subtraction problems. Bofferding and Richardson (2013) documented that many tended to use rules or order-based reasoning to solve problems, whereas Chiu (2001) found that students used multiple metaphors to explain how they obtained and justified their answers. However, the performance and reasoning of students who had only recently experienced integer instruction contrast with the above findings in two ways. First, Murray (1985) and Lamb et al. (2017) found that students were much less successful solving subtraction problems (39%–69% correct across both studies) than addition problems (about 75% correct). Second, in interviews with high-performing ninth-grade students, Murray (1985) found that errors were often due to misapplied rules (see, also, Vlassis, 2002, 2008). Thus, researchers found that after recent school-based instruction, participants found integer subtraction problems more challenging than integer addition problems and that errors were often due to misapplied rules.

Research involving students preinstruction. Early research on students' reasoning prior to school-based instruction on integer addition and subtraction provides consistent insights (Human & Murray, 1987; Murray, 1985; Peled, 1991; Peled et al., 1989). These researchers found that some children could solve many types of integer addition and subtraction problems using relatively abstract approaches, termed "spontaneous non-concrete reasoning" (Human & Murray, 1987, p. 438) and "mental models" that "were quite abstract" (Peled et al., 1989, p. 109). Use of these approaches contradicted the researchers' expectations that students would make connections to concrete embodiments of negative numbers, such as temperatures (Human & Murray, 1987; Murray, 1985) or debt (Peled, 1991). Additionally, Murray (1985) identified three productive strategies that the students used: (a) motion on "a vertical number line," (b) "correspondence (analogies) with operations on whole numbers" (p, 149; also identified in Human & Murray, 1987), and (c) logic in comparing the relationship between a previously solved problem and a related new problem (e.g., 5 + -3 and 5 – -3) to aid in solving the latter problem (p. 150).

More than 20 years after these researchers reported promising initial findings about students' integer reasoning prior to school-based instruction, other researchers also found that some children were capable of reasoning about integers in relatively sophisticated ways (e.g., Behrend & Mohs, 2005; Bishop, Lamb, Philipp, Schappelle, & Whitacre, 2011; Bishop, Lamb, Philipp, Whitacre, & Schappelle, 2014; Bofferding, 2010). For example, Bishop, Lamb, Philipp, Schappelle, and Whitacre (2011) found that first graders who had never heard of negative numbers nonetheless began to invent and reason productively about them in the contexts of completing addition and subtraction open number sentences. Additionally, Bofferding (2010) investigated 22 second graders' meanings for the minus sign, prior to instruction about negative numbers. The children solved an average of 20% of integer addition and subtraction problems correctly, using multiple meanings of the minus sign including subtraction, negation, and the sign of a number.

Across this body of studies, researchers found that some students could solve a range of addition and subtraction problems using relatively sophisticated and abstract strategies and that students used three productive ways of reasoning: (a) leveraging order relations (Bishop, Lamb, Philipp, Whitacre, & Schappelle, 2014; Bishop, Lamb, Philipp, Whitacre, Schappelle, & Lewis, 2014; Bofferding & Richardson, 2013; Chiu, 2001; Peled, 1991), (b) applying magnitude-based reasoning (Bishop, Lamb, Philipp, Whitacre, Schappelle, & Lewis, 2014; Chiu, 2001; Human & Murray, 1987; Whitacre et al., 2012), and (c) using logical necessity (Bishop, Lamb, Philipp, Whitacre, Schappelle, & Lewis, 2014; Bishop, Lamb, Philipp, Whitacre, & Schappelle, 2016a, 2016b; Murray, 1985). Murray's (1985) findings that students used three productive strategies (number line, analogy to whole numbers, and logic) closely correspond to three of the ways of reasoning identified by Bishop, Lamb, Philipp, Whitacre, Schappelle, and Lewis (2014): Motion on a vertical number line is an example of students' leveraging order relations; correspondence (analogies) to whole numbers is an example of magnitude-based reasoning; and employing logic aligns with logical necessity. The notable consistency between findings related to these ways of reasoning—shared in studies that were conducted 30 years apart and across different grade levels—provides strength for claims that these ways of reasoning are robust and consistently held by students.

Both Murray (1985) and Bishop, Lamb, Philipp, Whitacre, Schappelle, and Lewis (2014) identified important and consistent categories of student thinking, and Bishop, Lamb, Philipp, Whitacre, Schappelle, and Lewis (2014) provided an organizing framework (the Ways of Reasoning Framework) to structure the field's understanding and interpretation of student thinking with respect to integers. However, this framework has not been applied to a large data set. In this article, we present results of students' reasoning across a large, cross-sectional data set and demonstrate the Ways of Reasoning Framework's potential to unify students' integer thinking across grade levels.[2] We also add to the literature on problem types and flexibility by relating these constructs to integer operations of addition and subtraction. In the next section, we share research about problem types and how problem types may influence performance and strategy selection.

[2] We introduce this framework in more detail in the Conceptual Framework section.

Problem Types

Researchers have long recognized the importance of distinguishing among problem types within particular content areas and number domains (Carney, Smith, Hughes, Brendefur, & Crawford, 2016; Carpenter et al., 2014; Carpenter & Moser, 1982, 1983; Taber, 2002; Vergnaud, 1982, 1983). Features used to classify problems have included number choice (Carney et al., 2016), number relationships within problems (Taber, 2002), location of the unknown (Carpenter et al., 2014; Carpenter & Moser, 1982), operation (Vergnaud, 1983; Carpenter & Moser, 1983), and mathematical structure (Vergnaud, 1982, 1983). In most cases, researchers classified problems because of their influence on students' success rates (e.g., Carpenter & Moser, 1982, 1983; Taber, 2002) or children's problem-solving approaches (e.g., Carney et al., 2016; Carpenter & Moser, 1982). Because our ultimate goal is to share problem types that have the potential to support students' growth in mathematical understanding, we, too, focus on features of problems that distinguish differences in students' success and problem-solving approaches.

Within integer addition and subtraction, some work has been done to distinguish problem types (Bofferding, 2010; Bofferding & Richardson, 2013; Wessman-Enzinger, 2015). For example, Wessman-Enzinger (2015) identified 24 problem types, taking into account operation; sign; location; and magnitude of each addend, minuend, or subtrahend (e.g., $-4 + 6$ and $6 + -4$ are different problem types, which also differ from $-4 + 2$, and $-4 + 4$). Similarly, Bofferding (2010) shared 32 problem types, accounting for "the position of the signs and numerals in the problems" (p. 705). In subsequent work, Bofferding and Richardson (2013) identified problem types distinguished by operation and the number of positives or negatives in the problem statement. They shared six categories: (a) Addition: Two Positives; (b) Addition: Two Negatives; (c) Addition: Negative, Positive; (d) Subtraction: Two Positives; (e) Subtraction: Two Negatives; (f) Subtraction: Negative, Positive. Human and Murray (1987) shared 10 "cases" (p. 440) of result-unknown integer addition and subtraction problems, although they acknowledged that their list was incomplete (e.g., the cases did not include addition problems with the first addend negative and second addend positive). And we infer from Peled's Levels of Knowledge (1991), although not stated explicitly, that she distinguished three classes, or types, of problems on the basis of students' abilities to solve result-unknown integer addition and subtraction problems:

(a) $a \pm b$, for a an integer and b a natural number;
(b) $-a + -b$ for a, b natural numbers, $-a - -b$ for a, b natural numbers AND $a > b$; and
(c) $a \pm -b$ for a, b natural numbers, $-a - -b$ for a, b natural numbers AND $a < b$.

In the realm of integers, researchers have determined problem types on the basis of operations (e.g., Bofferding, 2010; Bofferding & Richardson, 2013; Murray, 1985; Wessman-Enzinger, 2015); whether the addends, minuend, or subtrahend were positive or negative (e.g., Bofferding & Richardson, 2013); how students solved the problems (e.g., Peled, 1991); and success rates (e.g., Murray, 1985).

Given the range in the number of problem types identified by different researchers, the field does not appear to have an established organizing principle to drive the identification of problem types. Thus, developing a coherent classification of problem types would address a gap in the literature and make a significant contribution to research in this area.

In our interviews with students, we found that some problems seemed to elicit particular approaches for solving them. For example, we found that some participants tended to approach the problem -3 + 6 = ☐ by counting on from -3 or using a number line. However, participants tended to approach the problem -5 + -1 = ☐ by comparing negative numbers to natural numbers. We thus sought to develop a problem-types framework for integer addition and subtraction that could be used to distinguish "important differences in how children solve the different problems" (Carpenter & Moser, 1983, p. 15).

In the next section, we share research about a related topic, flexibility, and how the construct of flexibility may play out for students' completion of integer open number sentences.

Flexibility

In his discussion of strategies for integer computation, Murray (1985) observed that the same child might use different strategies depending on the problem structure:

> Students were willing to change their strategies to accommodate the different cases, e.g. [*sic*] starting off with a vertical number line to deal with 5 – 8, but solving cases like -5 – -2 or -4 x 5 by extrapolating from known number facts. (p. 149)

During our interviews, we also noticed that some students seemed to purposefully and effectively change strategies from problem to problem. We were curious about the role that flexible strategy use, which we call *flexibility*, might play in relation to students' success in solving integer problems.

In the 1990s, several researchers documented that experts demonstrated flexibility (sometimes called *within-individual variability*, e.g., Dowker, Flood, Griffiths, Harriss, & Hook, 1996) when completing tasks in content areas such as single-digit multiplication (LeFevre et al., 1996), fraction comparison (Smith, 1995), and estimation (Dowker et al., 1996; Sowder, 1992). For example, Smith (1995) studied 11 competent students' strategies for solving a variety of rational number problems. On comparison tasks, he found that the students used different strategies from problem to problem and that the problem type was a better predictor of strategy use than the individual student. Smith shared, "These results indicate that these students selected strategies to take advantage of the possibilities presented by the numerical features of the items to achieve solutions with a minimum of arithmetical computation 'overhead'" (p. 28).

In recent studies, researchers have begun to focus more centrally on characterizing and operationalizing features of flexibility (Berk, Taber, Gorowara, & Poetzl, 2009; Star & Newton, 2009; Star & Rittle-Johnson, 2008) and investigating how to improve students' flexibility (Berk et al., 2009; Blöte, Van der Burg, &

Klein, 2001; Star et al., 2015; Star & Rittle-Johnson, 2008; Star & Seifert, 2006). For example, Berk, Taber, Gorowara, and Poetzl (2009) documented prospective teachers' flexibility in relation to proportional reasoning and designed an intervention to improve flexibility. They found that prospective teachers exhibited limited flexibility prior to participating in a methods class but that they learned to become more flexible, even 6 months after the intervention.

Star and colleagues have successively refined definitions of flexibility over time when new studies yielded insight. Initially, they defined flexibility as knowledge of multiple solution procedures as well as the capacity to innovate to create new procedures (Star & Seifert, 2006). In later work, they added knowledge of the relative efficiency of the strategies in addition to just knowing multiple strategies (Star & Rittle-Johnson, 2008). And later still, they defined flexibility as "knowledge of multiple solutions [or strategies] as well as the ability and tendency to selectively choose the most appropriate ones for a given problem and a particular problem-solving goal" (Star & Newton, 2009, p. 558). The most recent definition includes knowledge of strategies and selectively choosing strategies (rather than knowing only which strategies were relatively efficient) with respect to a specific goal.

Given the importance of flexibility demonstrated across several content areas and our own observations that some students we interviewed appeared to change strategies from problem to problem, we wondered about the role that flexibility might play in students' solutions to integer open number sentences. For example, do students who use multiple strategies select different strategies on the basis of particular problem types, and to what degree does this kind of flexibility influence performance? To answer these questions about flexibility, we adopted Star and Newton's (2009) definition, in which they emphasize not only having multiple strategies available to use but also selectively using those strategies on the basis of problem features.

To summarize, our goals in this article are to describe patterns in students' reasoning about integer addition and subtraction within and across grade levels and to identify possible relationships among students' problem-solving strategies, problem types, and the flexibility of students' problem solving. Given these goals, in the next section we share a framework for ways of reasoning about integer addition and subtraction that we developed and shared in earlier work and that also builds on other scholars' research related to student thinking in the domain of integers. We use the Ways of Reasoning Framework as an organizing structure for the research that we share here.

Conceptual Framework

The Ways of Reasoning Framework. The Ways of Reasoning Framework (WoRs) was developed and refined iteratively over a period of 3 years, using a large cross-sectional data set, and earlier versions have been shared (e.g., Bishop, Lamb, Philipp, Whitacre, Schappelle, & Lewis, 2014; Bishop et al., 2016a; Bishop, Lamb, Philipp, Whitacre, & Schappelle, in press). We identified five broad categories that we call *Ways of Reasoning* and multiple strategies within each way of reasoning that include more detail and nuance regarding a child's reasoning. WoRs are general conceptualizations of and approaches to solving integer addition and subtraction problems, characterized on the bases of key features of students' solutions and the underlying views of number and operations at work. We identified five WoRs that students across all participant groups in our study used when completing open number sentences: order-based, analogy-based, formal, computational, and emergent. (In earlier publications, we referred to analogy-based as *magnitude* and emergent as *limited*.) In Table 1, we define each way of reasoning (WoR).

Within each WoR we identified specific strategies that students used (e.g., counting as a particular instantiation of an *order-based* WoR or the use of a number line as a different instantiation of *order-based* reasoning). For us, a *strategy* is a subcategory of a particular WoR that further describes and differentiates student responses within the broader WoR. We view the five WoRs as an organizing structure into which we have categorized more detailed strategies on the basis of the underlying views of number and operations leveraged in a given strategy's use. We now briefly share descriptions of the five WoRs along with some of the corresponding strategies. (For more detailed descriptions of the WoRs and corresponding strategies, see Bishop et al., in press.)

Order-based. In an order-based WoR, one leverages the sequential and ordered nature of numbers to reason about a problem. Common strategies include using a number line (or motion), counting by ones, and jumping to zero (when computing, using friendly number decompositions that have a sum or difference of 0). For example, a student who describes completing $-4 + 7 = \square$ by starting at -4, "jumping" 4 units to reach zero, and then moving 3 more units beyond zero is using the *jumping to zero* strategy.

Analogy-based. An analogy-based WoR is often tied to ideas about cardinality and magnitude and is characterized by creating an analogy between signed numbers and some other concept. For example, many students compare negative numbers to positive numbers, using a strategy that we called *negatives like positives* (see also Human & Murray, 1987; Murray, 1985). This strategy involves computing with negative numbers through explicit comparison to computing with positive numbers (e.g., completing $\square + -2 = -10$ by comparing it to the problem $\square + 2 = 10$).

Formal. In reasoning formally, one treats negative numbers as formal objects that exist in a mathematical system and are subject to fundamental mathematical principles that govern their behavior. Formal strategies included *Infers Sign* and *Logical Necessity*. Ann's response shared at the beginning of the article is an example of Infers Sign. She looked at structural features of the problem—the operation in conjunction with the sign of the given numbers—to determine the sign of the answer prior to determining the final answer. We considered Ann's strategy to be a formal WoR because she made a claim about a class of problems—

Table 1
Ways of Reasoning Framework

Ways of reasoning		Definition
Evidence of engagement with negative numbers	Order-based	In this way of reasoning, one leverages the sequential and ordered nature of numbers to reason about a problem. Strategies include use of the number line with motion as well as counting forward or backward by 1s or another incrementing amount.
	Analogy-based	This way of reasoning is characterized by relating numbers and, in particular, signed numbers to another idea, concept, or object and reasoning about negative numbers on the basis of behaviors observed in this other concept. At times, signed numbers may be related to contexts (e.g., debt or digging holes).[a] Analogy-based reasoning is often tied to ideas about cardinality and understanding a number as having magnitude.
	Formal	In this way of reasoning, signed numbers are treated as formal objects that exist in a system and are subject to mathematical principles that govern behavior. Students may leverage the ideas of structural similarity, well-defined expressions, the structure of our number system, and fundamental principles (such as the field properties). This way of reasoning includes generalizing beyond a specific case by making a comparison to another, known, problem and appropriately adjusting one's heuristic so that the logic of the approach remains consistent, or generalizing beyond a specific case to apply properties of classes of numbers, such as generalizations about zero.
	Computational	In this way of reasoning, one uses a procedure, rule, or calculation to arrive at an answer. For example, some students used a rule to change the operation of a given problem along with the corresponding sign of the subtrahend or second addend (i.e., changing 6 – -2 to 6 + 2 or 5 + -7 to 5 – 7). Students often explained these changes by referring to rules like "keep change change" (***keep*** the sign of the first quantity, ***change*** the operation, and ***change*** the sign of the second quantity). For a strategy to be placed into this category, the student may state a procedure or rule with or without sharing a justification.
Restricted to whole numbers	Emergent	This category of reasoning reflects preliminary attempts to compute with signed numbers; the domain of possible solutions appears to be restricted to whole numbers. The effect (or possible effect) of operating with a negative number is not considered. For example, a child may overgeneralize that addition always makes larger and, as a result, claim that a problem for which the sum is less than one of the addends (6 + ☐ = 4) has no answer.

[a] If students related signed numbers to an order-based context such as elevation, they received both an order-based and an analogy-based code. Of the 115 contexts shared, only 6% (7) were order-based; all used the context of digging and filling holes.

addition problems wherein the sum is smaller than the addend—rather than referencing the specific values in this problem to initially determine the sign of the answer.

Computational. When students invoked a computational WoR, they typically used a procedure, rule, or calculation to arrive at an answer. See the example in Table 1.

Emergent. Strategies within the emergent WoR were locally restricted to whole numbers, and thus one would have to emerge from using this WoR to another to meaningfully compute with signed numbers (see also Whitacre et al., 2016). We view these strategies as sensible and reasonable, particularly given young children's early experiences with number. The most common strategy within the emergent WoR was *Addition Makes Larger/Subtraction Makes Smaller (AML/SMS)*, which is related to conceptualizations of addition as increasing the cardinality of a set and subtraction as decreasing the cardinality of a set; these stem from the overgeneralizations that addition always makes larger and subtraction always makes smaller. As an example of AML/SMS, consider Sam's response to $6 + \square = 4$, shared at the beginning of the article.

We use the five WoRs described above along with their corresponding strategies as a conceptual framework to help us identify important differences and similarities among students' integer reasoning and to describe the complexity and richness of students' thinking about integer addition and subtraction. The organization of strategies into broader WoRs within our framework enables us to distinguish key details of student thinking in a way that provides an overarching structure and coherence and, at the same time, to recognize important similarities in problem-solving strategies. Thus, we use the WoRs Framework as both a conceptual tool and an analytic tool in our study to help us identify patterns in students' reasoning, to consider different problem-type categories and how different categorizations may be related to students' WoRs, and to document problem-solving approaches from which flexibility may be measured.

Questions Stemming From Existing Research

As described in the previous sections, existing research about integer addition and subtraction indicates that (a) college students successfully solve problems using rules and order-based strategies and can explain their strategies using metaphorical reasoning; (b) after integers instruction, middle and high school students struggle with integer subtraction; (c) some young children can solve problems and reason productively about integers prior to school-based instruction; (d) problem types for integer addition and subtraction have been shared using a variety of classifications, identifying from two to 32 problem types; and (e) students appear, at times, to flexibly choose particular problem-solving strategies about integers on the basis of the problem features. As mentioned, we found no study in which elementary, middle, and high school students were systematically sampled to (a) document their WoRs, (b) identify problem types that evoke particular WoRs, or (c) examine the degree of flexibility in students' reasoning about integer addition and subtraction. We thus posed the following research questions to help address these gaps in existing scholarship:

1. How successful are students at solving integer addition and subtraction open-number-sentence problems, and what WoRs do they use to solve them?
2. What problem types, if any, are likely to evoke particular WoRs?
3. To what degree do students flexibly use analogy-based, computational, formal, and order-based WoRs, and what is the relationship between flexibility and students' success in completing the open number sentences?

Methods

Setting and Participants

We designed, conducted, and analyzed 160 interviews across four groups of students. We sought to map the terrain of endpoints for integers conceptions for K–12 students. At one end point, we included 40 students in Grade 11 whom we deemed to be mathematically successful and college-track, as determined by their enrollment in precalculus or calculus. To map the other endpoint for integers conceptions for K–12 students, we included 41 students in Grades 2 and 4 who had not yet heard of negative numbers. To complete the terrain, we also included a group of 39 students in Grades 2 and 4 who had heard of negative numbers but had not yet received school-based instruction about negative numbers and another group of 40 students in seventh grade who had already experienced school-based instruction with negative numbers. Our decision to group second- and fourth-grade students together on the basis of their familiarity with negative numbers resulted from discussions about the theoretical and practical benefits and drawbacks of grouping by grade level versus grouping by familiarity with negative numbers. We decided on these groupings because we deemed that the types of responses second and fourth graders who had (or had not) heard of negative numbers were more similar to each other, given our project goals. Moreover, we believed that the analyses would be more meaningful when comparing groups of students who had negative numbers in their mathematical domains versus those who did not than comparing students across grade levels (i.e., comparing second and fourth graders to each other). Further, when we took grade level into account in our within-group problem-type analyses, we found no statistically significant differences between second-grade and fourth-grade students.

The 160 students in the study were from 11 public schools (three elementary schools, three middle schools, one K–8 school, and four high schools) with varying standardized test scores as indicated by each school's Academic Performance Index (API). We purposefully selected schools to ensure a range of demographics and performance markers (Patton, 2015). For each grade level, we selected schools that represented a range of API scores, socioeconomic status (as determined by percentage of students on free or reduced-price lunch), and diversity. School size, API, and demographic information are provided in Table 2.

At each participating school, two teachers at each targeted grade level volunteered to participate. All students in the teachers' classes were invited to participate. Students were then selected randomly from among all who returned signed consent forms. We selected nine to 11 students per grade level from each of the four schools associated with each grade level. Our goal was not to relate specific instructional experiences to students' conceptions. Rather, we sought to understand the range of conceptions observed and the frequencies of these conceptions.

We named the groups according to the rationale for selection in this study: College-Track students (CTs; n = 40, 11th graders enrolled in precalculus or calculus), Post Instruction With Negatives students (PINs; n = 40, seventh-grade students who had recently completed instruction in integers), Before Instruction With Negatives students (BINs; n = 39, second and fourth graders with negatives in their numerical domains), and No Evidence of Negatives students (NENs; n = 41, second and fourth graders without negatives in their numerical domains). Group placements for second and fourth graders were made on the basis of responses to Questions 2, 3, and 9 in the interview (see the Introductory Questions section of the interview in the Appendix for Questions 2 and 3). The BIN group included 13 Grade 2 students and 26 Grade 4 students who, on the basis of responses to those two tasks, provided evidence of having at least some knowledge of negative numbers. The NEN group included 27 Grade 2 and 14 Grade 4 students who, on the basis of responses to those two tasks, provided no evidence of having knowledge of negative numbers. Students in the NEN group completed a subset of the interview items, given their lack of familiarity with negative numbers.

Interviews

The interviews, conducted at the children's school sites during the school day, were videotaped and typically lasted 60–90 minutes. The interviews consisted of a range of tasks, including open number sentences, number comparisons, and story problems. The interview protocol was developed iteratively while we tested tasks during pilot interviews and modified the tasks and sequence. A total of 74 pilot interviews were conducted in 2010. The final version of the interview included 56 items. In this article, we include analyses of the open number sentences portions of the interviews (shown in the Appendix).

When we initially designed the open number sentences, we considered problem features highlighted in the literature, such as the operation, a variety of sign combinations in each open number sentence, and the magnitudes of the numbers. In the final interview shared in the Appendix, we also added a dimension not included in the literature: We varied the location of the unknown (start, change, and result) with an emphasis on change- and result-unknown problems because of their potential to highlight differences in students' approaches.

Interviewers followed an interview protocol using a standardized open-ended interview (Patton, 2015). We asked students to read the question aloud, solve the

Table 2

School Demographic Characteristics

School	Grades interviewed	Number of students	Academic Performance Index (API)[a] Accountability Rating	Percentage Hispanic	Percentage White	Percentage African American	Percentage Asian	Percentage economically disadvantaged
School A	2 & 4	408	727	60%	2%	13%	20%	100%
School B	2 & 4	373	893	25%	63%	3%	1%	21%
School C	2 & 4	420	765	26%	54%	7%	2%	66%
School D	2, 4, & 7	619	862	51%	23%	17%	2%	52%
School E	7	1,011	690	76%	1%	9%	12%	100%
School F	7	593	763	50%	20%	12%	8%	78%
School G	7	1,161	836	32%	53%	5%	2%	30%
School H	11	1,847	756	86%	5%	4%	1%	72%
School I	11	1,726	786	40%	24%	19%	5%	67%
School J	11	1,164	742	55%	21%	15%	1%	63%
School K	11	1,423	807	15%	77%	3%	1%	18%

[a] API scores are based on the results of standardized tests administered annually to students in Grades 2–12. The API is a single score that combines students' scores on assessments across multiple content areas and ranges from a minimum of 200 to a maximum of 1,000. The state API average across all grade levels was 790 at the time of the interviews.

problem, and then immediately explain how they thought about the problem. Follow-up questions were restricted to seeking clarification on the strategy so that the interviewer understood the child's approach, except on the first open number sentence, 3 – 5 = □, when we asked specific follow-up questions on the basis of students' responses to confirm whether students had heard of negative numbers.

Analyses

In our analyses, we explore students' WoRs across participant groups, identify integer addition- and subtraction-problem types, and describe the relationships between problem types and WoRs. Additionally, we computed a flexibility measure for every student to examine the relationship between flexibility and performance on the open number sentences. Below we provide details of the analyses.

Performance and WoRs on open number sentences. Interview data were coded at the problem level; every open number sentence was coded for the underlying way (or ways) of reasoning the student used, any strategies used, and correctness. We coded directly from the videorecordings of the interviews and referred to student work collected during the interview as needed. (See Table 1 for definitions and examples. The complete WoRs coding scheme is available at http://www.sci.sdsu.edu/CRMSE/projectz/movies.html.)

Reliability. The two primary coders trained six additional people to code student interviews. Twenty percent of the interviews were double-coded to ensure reliability in our code interpretations and to guard against coding drift, and random assignment was used to select interviews to be double-coded. Coders were blind to which interviews were double-coded. Interrater reliability was calculated at both the way-of-reasoning and strategy-code levels. Interrater agreement was 92% for way of reasoning and 83% for strategy codes. In the interest of reliability, additional interviews identified as challenging to code were also double-coded. In total, 42 (26.25%) of the 160 interviews were double-coded. All discrepancies were resolved through discussion.

Problem-type categories. As noted in the description of the interview, open number sentences were systematically chosen to include different problem features, and those features were deliberately varied across the collection of 25 open number sentences. For example, although we initially wondered whether addition problems would be easier than subtraction problems or whether the location of the unknown would influence success, we found that, across the data set, these features had less influence on success than the signs of the numbers and the locations of negative integers. Thus, the features we considered in the design of the open number sentences did not yield problem-type categories that reflected differences in students' success or solution processes.

During our data collection, we noticed that some problems tended to evoke some WoRs more than others. We sought to test this hypothesis as part of our data analysis. We thus ran two types of statistical analyses to explore the three problem-type categories—change-positive, all-negatives, and counterintuitive (described in the Findings section)—that emerged. First, we investigated which problem types were easier and harder for students (controlling for possible grade-level effects) using three 2 × 2 ANOVAs to test for differences in percentage correct across problem-type categories and whether percentage correct varied by grade. In each ANOVA, the first factor occurred within participants (pairwise differences in percentage correct for problem types), and the second factor occurred between participants (grade level).[3] The two factors for each of these ANOVAs were (a) pairwise differences in percentage correct for problem type and (b) grade level. The first ANOVA was a 2 (Problem Type: Change-positive percentage correct vs. All-negatives percentage correct) × 2 (Grade level: second vs. fourth), the second ANOVA was a 2 (Problem Type: Change-positive percentage correct vs. Counterintuitive percentage correct) × 2 (Grade level: second vs. fourth), and the third ANOVA was a 2 (Problem Type: All-negatives percentage correct vs. Counterintuitive percentage correct) × 2 (Grade level: second vs. fourth).

Second, we investigated differences in the use of a given WoR across problem type (controlling for possible grade-level effects), for example, in the use of order-based reasoning across the three problem types and by grade level. Thus, within each WoR, we examined the difference in percentage use for problem type and whether the percentage use was dependent on grade. For each of the five WoRs, we used three 2 × 2 ANOVAs (for a total of 15). The two factors for each ANOVA were (a) pairwise differences in percentage use for problem type and (b) grade level. For example, within the order-based WoR, we ran three ANOVAs: The first was a 2 (Problem Type: Percentage use of order-based WoR on Change-positive vs. Percentage use of order-based WoR on All-negatives) × 2 (Grade level: second vs. fourth) ANOVA; the second was a 2 (Problem Type: Percentage use of order-based WoR on Change-positive vs. Percentage use of order-based WoR on Counterintuitive) × 2 (Grade level: second vs. fourth) ANOVA; and the third was a 2 (Problem Type: Percentage use of order-based WoR on All-negatives vs. Percentage use of order-based WoR on Counterintuitive) × 2 (Grade level: second vs. fourth) ANOVA. The first factor occurred within participants (pairwise differences in percentage use for problem types), and the second factor occurred between participants (grade level). Because we used multiple ANOVAs in our analysis, we used the Bonferroni Adjustment to reduce the likelihood of a Type I error at the .05 level (statistically significant F-values occurred at a = .05/18 = .0028). Additionally, all pairwise comparisons were two-tailed.

Flexibility. We remind the reader that we adopted Star and Newton's (2009) definition of flexibility as "knowledge of multiple solutions [or strategies] as well as the ability and tendency to selectively choose the most appropriate ones for a

[3] As explained in detail in the Findings section, we restricted our analysis to BIN students and thus had only two levels for grade: second grade and fourth grade.

given problem and a particular problem-solving goal" (p. 558). To use this definition, we calculated a flexibility measure for each student by first identifying the number of times each way of reasoning was used on problems for which that WoR was aligned on the basis of the problem type. Because determining the alignment of a student's solution depends on the problem-type category, we share more specific details about how we measured flexibility in the Findings section after we have discussed our problem types.

In general, for each open number sentence, we identified the WoRs that were aligned with that problem. To determine whether a student had a "tendency to selectively choose" a WoR, we counted instances of use of a WoR only if the student used an aligned WoR three or more times for a given subset of problems. We selected this threshold because the use of a particular way of reasoning at least three times on separate open number sentences indicated that a student's use of a way of reasoning was not anomalous but, rather, selectively chosen. Moreover, because we restricted our analyses to whether participants used each WoR on a subset of problems according to the problem-type framework, the problems themselves set a de facto upper bound on the frequency with which we expected to observe different WoRs. Thus, our measure of flexibility captured the number of WoRs a student selectively used. For example, a student who used order-based reasoning on seven, analogy-based reasoning on one, formal reasoning on two, and computational reasoning on five open number sentences that were aligned with the problem types would receive a flexibility score of 2 for selectively using two WoRs (order-based and computational). We then used the flexibility score to explore the relationship between flexibility and performance on the 25 open number sentences using correlation coefficients.

Findings

In the following sections, we share findings in four groupings: (a) students' performance on the open number sentences, (b) students' use of the five WoRs, (c) three problem types that emerged, and (d) the degree of students' flexible use of the WoRs and its relationship to their success on the open number sentences.

Participant-Group Performance

As one might expect, the participant groups performed differentially on the open number sentences. We found that half of all students who had not yet received school-based instruction had heard of negative integers. The BIN students solved, on average, more than one third of the problems correctly ($\bar{x}$ = 35.3%). Thus, some young children reasoned productively about negative numbers, and this reasoning often led to correct answers. In contrast, the NEN students correctly solved only 6% of the 12 problems posed to them ($\bar{x}$ = 5.9%). Additionally, the PIN students solved about three fourths of the problems correctly ($\bar{x}$ = 73.3%). whereas CT students solved virtually all problems correctly ($\bar{x}$ = 98.2%). That is, we saw a ceiling effect for our CT students.

To understand the spread of the data, we examined the percentage of students in each participant group who solved less than 10% or more than 90% of all problems correctly and then examined the percentage correct in 20% increments (see Table 3). Students in the NEN and CT student groups had the most clustered sets of scores. Three fourths of the NEN group solved fewer than 10% of the problems correctly, and more than half were incorrect on every problem. In contrast, all but one of the CT students solved more than 90% of the problems correctly, including 33 of the 40 students who answered every problem correctly. Most of the BIN students solved one third to one half of the problems correctly. The PIN students' performance varied the most, with scores ranging from 16% to 100% correct. Two subgroups of PIN students emerged: a group correct on 90% or more of the tasks and a larger group of students (almost half) who failed to reach what many consider to be a minimum proficiency level of 70% correct.

We make three observations based on these data. First, the initial relative success of BIN students (given their lack of school-based instruction on the topic) may indicate that some students have fruitful conceptions on which teachers might build. Second, the PIN students had completed all school-based instruction on integers identified in the state standards at the time of the study, and although more than one third of the PIN students (37.5%) correctly answered 90% or more of the problems, half responded correctly to fewer than 70% of the items. Finally, because we intentionally selected CT students who were on a college-track trajectory, the CT students performed as expected: 39 of 40 CT students scored above 90%, and more than 80% answered every problem correctly.

Additionally, consistent with findings from Murray (1985) and Lamb et al. (2017), the PIN students, who had completed school-based instruction on integers within 2 years of the interviews, differed in their performance on subtraction problems compared with addition problems (means of 64% versus 84%, respectively). However, the other three groups, NEN, BIN, and CT, performed similarly on subtraction and addition problems (NEN: 6% correct for both; BIN: 36% vs. 34%, respectively, and CT: 98% vs. 99%, respectively). In the next section, we examine the frequency of use of the WoRs.

Table 3
Participant-Group Performance on Open Number Sentences

Participant group	< 10% correct	10–29% correct	30–49% correct	50–69% correct	70–89% correct	≥ 90% correct
NEN *n* = 41	75.6% (*n* = 31)	24.4% (*n* = 10)				
BIN *n* = 39	2.6% (*n* = 1)	30.8% (*n* = 12)	59% (*n* = 23)	5.1% (*n* = 2)	2.6% (*n* = 1)	
PIN *n* = 40		5% (*n* = 2)	12.5% (*n* = 5)	30% (*n* = 12)	15% (*n* = 6)	37.5% (*n* = 15)
CT *n* = 40				2.5% (*n* = 1)		97.5% (*n* = 39)

Frequency of Use of Ways of Reasoning

In this section, we share our second main finding and report the frequencies of WoRs used across the data set and within participant groups (refer to Table 1 in the Conceptual Framework for descriptions of these categories). As seen in Table 4, students in the BIN, PIN, and CT participant groups used each of the five broad WoRs shared in the conceptual framework. Across all problems posed, computational reasoning was the most common, occurring in responses to about two of every five problems. Emergent and order-based WoRs were used on about one third and one fourth of the problems posed, respectively. Formal and analogy-based WoRs were less frequent overall, with each employed in responses to about 10% of the problems posed.

Students in the NEN group provided no evidence of experience with negative numbers, and, unsurprisingly, they solved 90% of all problems with an emergent WoR. No problems for this participant group were solved using order-based, analogy-based, or formal WoRs. Although the BIN students solved more than half of the problems using an emergent WoR, they used all other WoRs as well: They solved one third of all problems with order-based reasoning, about 10% each with analogy-based and computational reasoning, and 3% with formal reasoning. The PIN group solved half of the problems with a computational WoR, almost 40% with an order-based WoR, and 13–20% with analogy-based, formal, or emergent WoRs. Finally, the CT group solved three fourths of the problems using a computational WoR, almost none (fewer than 1%) with an emergent WoR, about one fourth with a formal WoR, and almost one fifth each with order-based and analogy-based reasoning. As described in the Methods section, students can use more than one WoR when solving a problem—as was done by our CT students for many problems. The CT students used computational reasoning in conjunction with at least one additional WoR on 28% of all problems posed, and they solved 25% of the problems without a computational WoR. Thus, the CT students used a WoR *other than or in addition to* computation on more than half (53%) of the problems

Table 4
Frequency of Ways of Reasoning

	Way-of-reasoning percentage use (By total number of problems)						
Participant group	Order-based	Analogy-based	Formal	Computational	Emergent	Other	Unclear
NEN	0%	0%	0%	12%	90%	4%	2%
BIN	33%	11%	3%	9%	58%	5%	2%
PIN	38%	20%	13%	51%	16%	1%	1%
CT	19%	16%	24%	75%	1%	<1%	<1%
Overall	26%	13%	12%	41%	34%	2%	1%

Note. Because students can use more than one way of reasoning to solve a problem, row-percentage sums are larger than 100%.

posed, confirming not only the expected vast use of the computational WoR but also showing considerable unexpected use of the other WoRs.

Additionally, we identified differences and similarities in the patterns of use of WoRs across participant groups (again, refer to Table 4). First, in order from NEN, BIN, PIN, and CT students, the use of the emergent WoR decreased and formal WoR increased. Second, the BIN, PIN, and CT groups all used order-based and analogy-based WoRs on one tenth to one third of the problems. Third, the number of problems solved using more than one productive WoR per problem increased by group, with PIN and CT students using more than one WoR per problem on about one fourth and one third of all problems posed, respectively.

Problem-Type Categories

We now turn to a discussion of problem types for integer addition and subtraction as our third main finding. We distinguished three categories of problems of the form $a \pm b = c$ by the signs of their values. Our analysis revealed that the problem types (a) differed in difficulty and (b) tended to evoke different WoRs.

Characterizing problem-type categories. One can use the flow chart (see Figure 1) to determine category placement of any open number sentence of the form $a \pm b = c$ (some describe *a*, *b*, and *c* as the start, change, and result values, respectively [Carpenter et al., 2014]). In the first problem category, *all-negatives*, all three values in the problem are negative (the start, change, and result are all negative, e.g., $-5 + \square = -8$). In the second category, *change-positive* problems, *b* (the change value) is positive. The problems $3 - \square = -6$ and $-2 + 8 = \square$ are both change-positive because the *b* values of 9 and 8, respectively, are positive.

In the third category, *counterintuitive problems*, *b* (the change value) is negative, but at least one of *a* (the start) and *c* (the result) is positive.[4] For example, $6 + \square = 4$ and $-3 - \square = 2$ are counterintuitive problems because the *b* (change) values are negative, but the problems have at least one positive value. We named this class of problems counterintuitive because these problems contradict the overgeneralization almost all young children make that addition makes larger or that subtraction makes smaller. Thus, the problems are counterintuitive for students who are attempting to solve these problems for the first time (Bishop et al., 2011).

The three categories of open number sentences, examples of problems we posed within each category, and the average percentage correct for each category are shown in Table 5. Our analysis revealed that some problem types were more challenging than others. The overall percentages correct (aggregated across participant groups) show that the easiest problems were the *all-negatives* (76.3% correct), followed by *change-positive* (61.1% correct), and *counterintuitive* (48.4% correct).

Why problem types matter: An analysis of BIN students. Although problem types have the potential to evoke particular WoRs for students in every participant

[4] Having at least one positive value is necessary to avoid being an all-negatives problem.

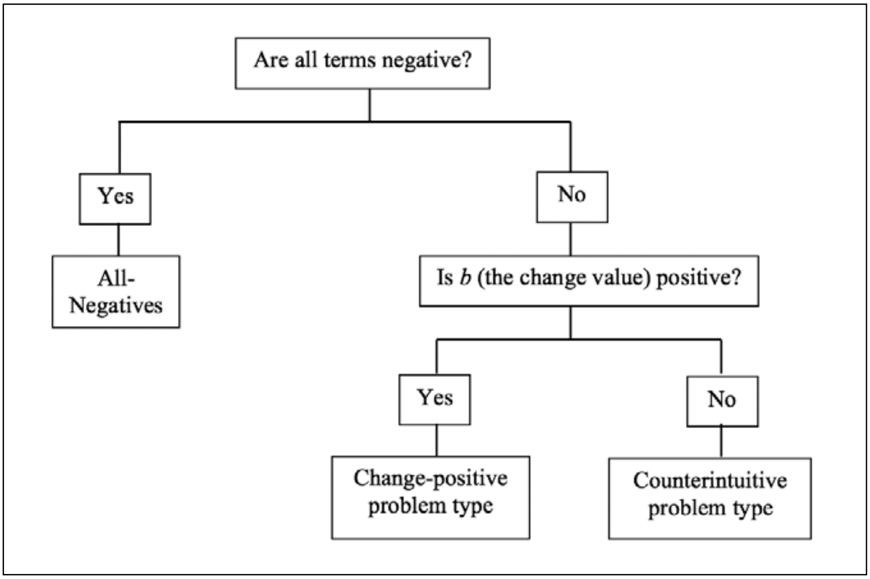

Figure 1. Flowchart for categorization of problems of the form $a \pm b = c$.

group, we focus on the BIN participant group. This group is particularly relevant for the problem-type analysis because students in this group had heard of negative numbers but had not received school-based instruction about negative numbers. Thus, this group (more so than the other groups) provides a window into students' initial ideas about solving problems with negative numbers. To support this claim, we consider the use of various procedures in the computational WoR for different participant groups. Whereas the PIN and CT students used procedures that are often taught in school (such as keep change change, equations, same-sign rule, and different-sign rule) 282 and 660 times, respectively, the BIN students used those same procedures only two times across all problems solved. Thus, we argue that because the BIN students did not use those commonly taught procedures, we had greater access to initial approaches for solving these problem types through them than through students in the other groups. Further, students using rules or procedures transformed problems to types different from the original problems. For example, a student who used the keep-change-change rule for the all-negatives problem $-5 - -3 = \square$ transformed the problem to $-5 + +3 = \square$, a change-positive problem. The student could then use various WoRs to solve the transformed problem (rather than the original problem), which would make linking the WoRs to the original form of the problem impossible. In contrast, the BIN students used almost no procedures that transformed the problems into different types, and thus their WoRs were directly related to the original versions of the problems.

Table 6 displays the mean percentage correct by problem type and the mean percentage use of a given WoR by problem type for BIN students. The overall

Table 5
Problem Types of the Form $a \pm b = c$

Category	Characterization	Sample problems	Average percentage correct
All-negatives	All negative values	□ + -2 = -10 -5 + □ = -8 -5 – -3 = □ -8 – □ = -2 -5 – -5 = □ [a]	76.30%
Change-positive	*b* (change) value positive	-3 + 6 = □ -2 + □ = 4 3 – □ = -6 -2 – 7 = □ -8 + □ = 0	61.10%
Counterintuitive	*b* (change) value negative, *a* (start), *c* (result), or both positive	6 + -3 = □ 6 + □ = 4 5 – □ = 8 6 – -2 = □ 3 + □ = 0	48.40%

[a] 0 may be thought of as both positive and negative, and so this problem was determined to be all negatives because the approaches BIN students took were more similar to those for all-negatives than for counterintuitive problems.

trend in percentage correct for the problem types is consistent with those for all participants in the study (that is, for the entire data set, all-negatives problems were the easiest for students to solve, followed by change-positive and counterintuitive). However, for BIN students, the differences in accuracy across problem types are noteworthy. Although BIN students correctly solved almost three fourths of all-negatives problems and one third of change-positive problems, they correctly solved fewer than one tenth of the counterintuitive problems. As a direct basis for comparing these two problem types, we asked the students to solve both the change-positive problem -3 + 6 = □ and the counterintuitive problem 6 + -3 = □.

From an expert perspective, these problems are equivalent (one needs only to invoke the commutative property of addition); however, from a young person's perspective, the problems differ. One half of the BIN students correctly solved the former, whereas only 13% correctly solved the latter.

As described in the Methods section, we used three 2 × 2 ANOVAs to test for differences in percentage correct across problem-type categories (controlling for grade level within the BIN group). We found that every percentage-correct pair-wise-comparison among the problem types was significant

$$[F_{\text{All-negatives vs. Change-positive}}(1, 37) = 24.86, p < .0001;$$
$$F_{\text{All-negatives vs. Counterintuitive}}(1, 37) = 124.06, p < .0001;$$
$$F_{\text{Counterintuitive vs. Change-positive}}(1, 37) = 43.99, p < .0001].$$

Table 6
Percentage Correct and Use of WoR by Problem Type for BIN Students

	Problem types of the form $a \pm b = c$		
	All-negatives	Change-positive	Counterintuitive
Percentage correct	74%	33%	9%
Percentage use of WoR			
Analogy-based	34%	5%	3%
Order-based	29%	44%	21%
Computational	8%	12%	7%
Formal	4%	4%	2%
Emergent	45 %	50%	77%

Additionally, we found no main effects for grade level. In other words, second graders did not differ from fourth graders in percentage correct overall or within the problem-type categories. The results of this analysis show that the three problem types differ in difficulty: All-negatives was the easiest type of problem, followed by change-positive, and counterintuitive was the most difficult problem type by a substantial margin.

We also conducted within-participant pairwise analyses of percentage use of each WoR among the three problem types.[5] For example, we tested to determine whether the fact that BIN students used analogy-based reasoning on 34% of the all-negatives problems but on only 3% of the counterintuitive problems was statistically significant. We present those findings in Table 7. Refer to the Methods section for additional details regarding the analyses. In the sections that follow, we describe the statistically significant problem-type comparisons and students' WoRs to solve them.

Problem-type comparisons within analogy-based WoRs. We found statistically significant pairwise differences in the percentage use of analogy-based reasoning between all-negatives problems and each of the other two problem types. Moreover, we found no significant differences in the use of the analogy-based WoR between change-positive and counterintuitive problem types. In other words, BIN students were more likely to use analogy-based reasoning on an all-negatives problem than on one of the other problem types. These results indicate that structural features of the all-negatives problems may have evoked analogy-based reasoning. We conjecture that problems with all negative values, as in $-5 - -3 = \square$, are easier than other problem types to compare to a related problem involving only natural numbers (i.e., $5 - 3 = \square$) or to think about the given problem in terms of objects or contexts because all quantities are of the same sort. With an all-

[5] We covaried any potential differences that may exist between the second and fourth graders.

Table 7
Pairwise Comparisons for Problem-Type Differences Within Each WoR (BIN students)

Way of reasoning	Problem type	Statistics
Analogy-based	All-negatives vs. Change-positive	$F(1,37) = 24.69, p < .0001$*
Analogy-based	All-negatives vs. Counterintuitive	$F(1,37) = 31.21, p < .0001$*
Analogy-based	Change-positive vs. Counterintuitive	$F(1,37) = 2.48, p > .12$
Order-based	All-negatives vs. Change-positive	$F(1,37) = 35.04, p < .0001$*
Order-based	All-negatives vs. Counterintuitive	$F(1,37) = 4.93, p < .05$
Order-based	Change-positive vs. Counterintuitive	$F(1,37) = 78.99, p < .0001$*
Computational	All-negatives vs. Change-positive	$F(1,37) = 1.75, p > .19$
Computational	All-negatives vs. Counterintuitive	$F(1,37) = 0.60, p > .40$
Computational	Change-positive vs. Counterintuitive	$F(1,37) = 5.36, p < .03$
Formal	All-negatives vs. Change-positive	$F(1,37) = 0$
Formal	All-negatives vs. Counterintuitive	$F(1,37) = 2.24, p > .14$
Formal	Change-positive vs. Counterintuitive	$F(1,37) = 3.21, p > .08$
Emergent	All-negatives vs. Change-positive	$F(1,37) = 0.19, p > .60$
Emergent	All-negatives vs. Counterintuitive	$F(1,37) = 33.31, p < .0001$*
Emergent	Change-positive vs. Counterintuitive	$F(1,37) = 78.90, p < .0001$*

*Statistically significant differences when test was conducted using the Bonferroni Adjustment at $\alpha = .0028$ (the total number of pairwise comparisons was 18, including both percentage correct and percentage use). Each comparison was two-tailed.

negatives problem, students are not forced to explicitly consider how opposite signs interact or even how to appropriately represent different signs. Instead, students can extend their understanding of addition and subtraction on natural numbers to the entire set of integers. Thus, all-negatives problems such as $-5 + -1 = \square$ and $-8 - \square = -2$ seem to lend themselves to students using an analogy-based-reasoning strategy such as negatives like positives, described earlier. As a basis for comparison, students solved both $-5 - -3 = \square$ (all-negatives) and $6 - -2 = \square$ (counterintuitive). Although both problems involve subtraction, have an unknown result, and involve a negative change value, 15 BIN students (38%) used an analogy-based approach for the first, whereas none used an analogy-based approach for the second.

Problem-type comparisons within order-based reasoning. Students used order-based reasoning on almost one half (44.29%) of the change-positive problems but on only about one third and one fifth of the all-negatives and counterintuitive problems, respectively. The differences between the use of order-based reasoning on change-positive problems compared with its use on all-negatives problems and counterintuitive problems were statistically significant. We suspect that change-positive problems such as $-2 + \square = 4$ and $3 - \square = -6$ tended to evoke more order-based reasoning than all-negatives problems or counterintuitive problems because of their structure. Because *b* (the change value) is positive for this problem type, addition and subtraction can be thought of in terms of motion to the right and left (or up and down), respectively, on the number line or counting up and down. Thus, the behavior of change-positive problems is consistent with the addition makes larger (AML) and subtraction makes smaller (SMS) generalizations that students may have made about adding and subtracting natural numbers. Moreover, these generalizations hold for operations with negative numbers so long as *b* (the change value) is positive (which it is for this problem type). For example, students solved both 3 – ☐ = -6 (change-positive) and 5 – ☐ = 8 (counterintuitive). Although both problems are subtraction and have an unknown subtrahend, they are classified as different problem types, and the BIN students differed in their approaches. Whereas 77% (30 of the 39) of the BIN students used order-based reasoning to solve the first problem, only 8% (3 of the 39) used order-based reasoning to solve the second (instead, 90% [35 of the 39] used emergent reasoning). Problems that conformed to the overgeneralization that addition makes larger and subtraction makes smaller seemed to support students in successfully extending the order-based reasoning they used for natural numbers on these new types of problems. Many students were potentially able to solve change-positive problems by invoking their current views about the nature of addition and subtraction, successfully extending those ideas to negative numbers, so long as the change value was positive.

Problem-type comparisons within emergent WoRs. In contrast to change-positive problems, problems with negative *b*-values contradict the overgeneralizations that addition makes larger and subtraction makes smaller. Every problem in this category is counterintuitive in that for addition problems the result is smaller than the starting value and for subtraction problems the result is larger than the starting value (that is, addition makes smaller and subtraction makes larger). The challenges of confronting these contradictions may explain not only why counterintuitive problems such as 6 + -3 = ☐ and 5 – ☐ = 8 were the hardest questions for this participant group to solve (they completed an average of fewer than 10% correctly) but also why they evoked emergent reasoning significantly more often than all-negatives problems or change-positive problems (see relevant pairwise comparisons in Table 7). BIN students solved more than three fourths of the problems in the counterintuitive category using emergent reasoning, but they used emergent reasoning on one half or fewer of the problems in the change-positive and all-negatives categories. To illustrate this relationship, we asked students to solve 6 + ☐ = 4 (counterintuitive) and -2 + ☐ = 4 (change-positive). Although both problems are addition, have an unknown change value, and have the same addends (6 and -2), 37 of 39 BIN students (95%) used emergent reasoning to complete the first problem, and only 10 of 39 (26%) used emergent reasoning to complete the second problem. Instead, as one might expect, 23 of 39 (59%) used order-based reasoning to complete the change-positive problem of -2 + ☐ = 4.

In summary, we used multiple ANOVAs to investigate the relative difficulties of our problems by problem-type category as well as to determine differences in the use of a given WoR across problem types. We found that some problems were easier than others for students and that particular WoRs were used more frequently with specific problem types. For BIN students, all-negatives problems were the easiest problems, and analogy-based reasoning was more likely than other WoRs to be used to solve these types of problems. Change-positive problems were the next easiest for BIN students, and order-based reasoning was more likely than other WoRs to be used on them. Finally, counterintuitive problems were the most difficult, and emergent reasoning was the most likely WoR to be used to solve counterintuitive problems. Our results show that the problem-type categories that emerged from our analysis do differentiate both students' success in correctly solving these problems and their approaches to solving them.

Two types of change-positive problems. Because of the differential success and conceptual differences (from a student's perspective) in solving problems within the change-positive category, we further differentiated this category by examining those problems that cross zero and those that do not. For example, 3 – ☐ = -6 is a *cross-zero* problem because the starting and ending values are on opposite sides of zero, and, as a result, the solution involves *crossing zero*. In contrast, to solve *negative-side* problems, like -2 – 7 = ☐, one does not cross zero. For example, a common order-based approach for completing -2 – 7 = ☐ is to start at -2 on the number line, move left seven units, and end at -9, thus staying on the *negative side* of the number line. Like cross-zero problems, negative-side problems conform to the overgeneralizations that addition makes larger and subtraction makes smaller (e.g., for the negative-side problem -2 – 7 = ☐, the difference, -9, is smaller than the minuend, -2). Table 8 shows that although students used order-based reasoning more than any other productive WoR for both subcategories of change-positive problems, they were less successful on negative-side problems than on cross-zero problems: 42% correct versus 19% correct, respectively [$F(1,37) = 18.91, p < .0001$]. Further, students were more likely to use an emergent way of reasoning to solve negative-side problems than cross-zero problems [$F(1,37) = 28.12, p < .0001$].

Consider the cross-zero problem 3 – ☐ = -6 and the negative-side problem -2 – ☐ = -8. These problems are similar in that they have a positive but unknown *b*-value (change value), involve subtraction, have a result that is negative, and, from an expert perspective, conform to the notion that subtraction makes smaller. The problems are different in that the first problem crosses zero, whereas the second does not. Moreover, this pair of problems conforms to the relative-difficulty pattern for the cross-zero and negative-side problems: 56% of the BIN students correctly solved 3 – ☐ = -6, but only 21% correctly solved -2 – ☐ = -8. We conjecture that the cross-zero problem was solved correctly more often than the negative-side problem for the following reasons: Most students knew that negative numbers were smaller than positive numbers (Whitacre et al., 2017), and because cross-zero

Table 8
Ways of Reasoning Used by BIN Students on Change-Positive Problems

	Change-positive	
	Cross-zero	Negative-side
Percentage correct	41.76%	18.59%
Percentage use for WoR		
Order-based	47.62%	38.46%
Analogy-based	3.30%	8.33%
Computational	14.29%	8.97%
Formal	5.49%	1.28%
Emergent	41.03%	65.38%

problems start and end on opposite sides of the number line, students could readily identify that addition made the sum larger and subtraction decreased the starting value. In contrast, when solving negative-side problems, students had to contend with notions of *bigger* and *smaller* that may not have been evoked for the cross-zero problems. For example, on the negative-side problem -2 – ☐ = -8, 54% of the BIN students claimed that the problem had no possible answer. These students reasoned that subtraction should make smaller, but because they believed -8 was *greater than* -2, solving the problem was impossible for them. Although -8 is *less than* -2, -8 has a *greater magnitude* than -2 and thus presents special conceptual challenges for students (Bofferding, 2014; Whitacre et al., 2016). Although some students were able to invoke an order-based strategy and, for example, move six units from -2 to -8 on the number line to get an answer of 6, the negative-side problems were answered incorrectly more often than the cross-zero problems.

In summary, we identified three main problem types of differing difficulty that evoked or have the potential to evoke different WoRs. These problem types are identified by the signs of the values in the problems and the locations of signed values. In the next section, we extend our work on problem types and examine the degree to which students flexibly reasoned about the open number sentences.

Flexibility

For our analysis, recall that we used Star and Newton's (2009) definition of flexibility: "knowledge of multiple solutions [or strategies] as well as the ability and tendency to selectively choose the most appropriate ones for a given problem and a particular problem-solving goal" (2009, p. 558). In this section, we describe both the degree of flexibility of students in the BIN, PIN, and CT participant groups[6] and the relationship between flexibility and accuracy when we share our fourth, and last, finding.

[6] Recall that participants in the NEN group provided no evidence of knowledge of negative numbers and that they provided 0 instances of the use of order-based, analogy-based, and formal reasoning. Thus, they are excluded from this analysis.

Table 9 displays the percentage of students in each participant group with flexibility scores of 0, 1, 2, 3, and 4. The maximum number of WoRs that a student could use is four because we excluded the emergent WoR (the only WoR wherein strategies were locally restricted to the domain of whole numbers) from this analysis. The values reflect the percentage of students who provided evidence that they selectively used a WoR (used a WoR three or more times) on problems for which the WoR was aligned with the problem type. For example, a student was determined to have selectively used an order-based way of reasoning if the student used that WoR on at least three change-positive problems. Similarly, a student had to have used analogy-based reasoning on at least three all-negatives problem[7] to have selectively used analogy-based reasoning, used computational reasoning at least three times on counterintuitive problems to have selectively used computational reasoning, and used formal reasoning at least three times on any problem to have selectively used formal reasoning, given that formal reasoning can be appropriately applied to all problem types. Computational reasoning was deemed a way of reasoning aligned with solving counterintuitive problems given that the use of a computation on problems of this type is an effective, efficient way to solve these problem types.

The BIN students selectively used, on average, 1.31 WoRs across all open number sentences posed in the interview, and almost all (87%) selectively used one or two WoRs. The PIN students selectively used, on average, 2.40 WoRs, and almost half (45%) selectively used three appropriate WoRs. Finally, the CT students selectively used, on average, 2.75 WoRs; half of them (48%) selectively used three appropriate WoRs, and about two thirds (68%) selectively used either three or four appropriate WoRs.

In Table 10, we further explore flexibility by sharing information about the percentage of appropriate problems for which students selectively used each WoR. Almost every PIN and CT student used a computational WoR on at least three problems for which its use would be considered aligned with the problem type,

Table 9
Flexibility by Group

Participant group	*n*	Flexibility score 0	1	2	3	4	Mean
BIN	39	10%	51%	36%	3%	0%	1.31
PIN	40	0%	15%	35%	45%	5%	2.40
CT	40	0%	13%	20%	48%	20%	2.75

[7] Because we identified that many PI and CT students often appropriately treated one meaning of the minus sign (i.e., binary operator [subtraction], unary operator [opposite of], or nonoperator [sign of the number]) as if it were another (Lamb et al., 2012), we also included in this count instances of analogy-based reasoning on the following problems: $-8 - 3 = \square$, $-2 - \square = -8$, and $-2 - 7 = \square$, problems for which some students claimed that two negatives were in the problem.

and 40% or more of them selectively used each of formal, analogy-based, and order-based WoRs. In contrast, about three fourths of the BIN students used order-based reasoning on at least three problems.

The relationship between flexibility and performance. Because of the relationships that we identified between WoRs and problem types, we wondered how WoRs and performance might be related. If some problem types lend themselves to particular WoRs, would those students who had multiple WoRs at their disposal and used them with particular problem types be better equipped to correctly solve problems? In investigating this relationship between percentage correct and flexibility, we found that flexibility was positively correlated with performance in our data, both across participant groups (r = .384, one-tailed, p < .01) and within participant groups (r = .277, .534, .345 for BIN, PIN, and CT, respectively, one-tailed, all p-values < .05). We wondered about potential alternative explanations for these correlations between flexibility and success. For example, might the PIN and CT students' increased use of computational and formal reasoning relative to the BIN students' use (see Table 4) be underlying mechanisms that account for the improved performance? However, we believe that the statistically significant correlations *within every participant group* (even those groups with and without high rates of the use of computational reasoning) provide sound evidence for our assumption that the overarching degree of flexibility, rather than an increased use of a particular WoR, explains improved student performance. That said, we are open to the possibility that mechanisms we did not measure may underlie students' flexibility and that such mechanisms could account for some students being both more flexible and more accurate than others.

In summary, we have documented statistically significant moderate-to-strong correlations across and within participant groups between flexibility and performance on integers open number sentences, and we found that the correlations were strongest in the PIN group.

Participant-Group Summaries

We now briefly summarize our findings related to each participant group. First, NEN students struggled to answer any problems; almost exclusively used emergent reasoning to solve them; and provided no instances of using order-based, analogy-based, or formal WoRs. Second, although BIN students were also likely to use emergent reasoning to solve problems, they also correctly solved about one third

Table 10
Percentage of Students Who Selectively Used Each WoR

Group	Order-based	Analogy-based	Computational	Formal
BIN	74%	41%	5%	10%
PIN	63%	40%	85%	53%
CT	40%	55%	100%	80%

of all problems, used order-based reasoning more than any of the other productive WoRs, and used all productive WoRs. They were not flexible in their problem-solving approaches, with more than half of the students selectively using only one WoR. Although many BIN students were able to solve all-negatives and change-positive problems, they correctly solved very few counterintuitive problems. Third, the PIN students had the greatest range in performance and the strongest correlation between flexible use of the WoRs and performance. They solved more than half of all problems using a computational WoR and used each of the other WoRs on between one eighth and one third of the problems. They were the only students who were much more successful in completing addition than subtraction open number sentences. Fourth, the CT students, selected to represent the upper range of expertise with respect to integer addition and subtraction at the high school level, answered almost every problem correctly, used computational reasoning more than any other productive WoR, almost never used emergent reasoning, and were the most flexible in their problem solving. The finding that CT students were both the most flexible in their WoRs and the most accurate indicates that particular WoRs are not necessarily replaced by other, more sophisticated WoRs.

Summary of Findings

In response to our research questions, we found that across participant groups, students used five broad WoRs: order-based, analogy-based, computational, formal, and emergent. In relation to the WoRs, we identified three major problem types: all-negatives, change-positive, and counterintuitive. The problem types are distinguished on the bases of the signs of the numbers and the locations of the signed numbers, and we found that they evoked or have the potential to evoke particular WoRs. Finally, we found that the students' average degree of flexibility increased in order of BIN, PIN, and CT students. Moreover, flexibility and performance were positively correlated both within and across participant groups. In the discussion, we share implications from these findings with respect to the teaching and learning of integers.

Discussion

Findings from our cross-sectional study indicate that the WoRs Framework can be applied to the integer reasoning of students in elementary, middle, and high school and that integers problem types may help teachers anticipate students' performance and approaches. Both the WoRs and the problem-types frameworks provide structures that teachers, researchers, and professional developers may be able to leverage when working to develop and investigate students' understanding of integer addition and subtraction. Additionally, we found that students have the capacity to grapple with sophisticated mathematical ideas when completing integers addition and subtraction open number sentences. We suggest that our study provides three types of resources for educators: (a) WoRs and problem-types frameworks, (b) characterization and development of flexibility, and (c) development of a trajectory of learning about integers.

Ways of Reasoning and Problem-Types Frameworks

Ways of Reasoning Framework. Teachers have a great deal of knowledge about their students' mathematical ideas, but without conceptual frameworks for interpreting those ideas, they are often unable to make use of this knowledge in their teaching. We know that organized conceptual frameworks for classifying students' mathematical ideas provide lenses for teachers and researchers to attune to students' ideas and develop a structure for assessing and supporting students' understanding (Carpenter, Fennema, Peterson, & Carey, 1988; Carpenter et al., 2014; Carpenter et al., 2003; Empson & Levi, 2011). The degree with which students in our cross-sectional investigation used the WoRs indicates that the framework can be used across grade levels K–12. Additionally, we have not seen many of the most productive strategies within the framework, such as treating negatives like positives, inferring the sign of the number prior to determining the final answer, or invoking logical necessity (Bishop et al., 2011, Bishop et al., 2016a) in textbooks or other curricular materials. Thus, access to the framework may provide a new way for teachers to classify the reasoning their students use so that they can build on that reasoning.

The role of problem types in supporting students' learning. Understanding how classes of problems may influence students' approaches may be critical knowledge for teachers and researchers to gain to support and investigate students' learning. In our work, we identified three problem types—change positive, all negatives, and counterintuitive—that have the potential to evoke particular WoRs and that are differentially challenging for students to solve. When teachers and researchers are equipped with a structure for organizing both WoRs and problem types that may evoke particular WoRs, they may be able to provide specific, timely, research-based support to students (cf. Carpenter et al., 2014). Additionally, this study is the first in which change- and start-unknown problems were used to investigate integer addition and subtraction (cf. Vlassis, 2002, 2008, who investigated change-unknown problems in the context of solving algebraic equations). In particular, counterintuitive change-unknown problems such as $6 + \square = 4$ and $5 - \square = 8$ explicitly highlight the common AML/SMS overgeneralization. Prior to instruction, most students will claim that these problems cannot be solved. These problems can serve a pedagogical role by focusing students' attention on this conundrum: Is it possible to add and arrive at an answer that is less than the starting value, and is it possible to subtract and get a difference larger than the minuend? These particular types of open number sentences may play a pivotal role in supporting the extension of students' numerical domains to permit negative numbers to serve as change values in open number sentences.

Acquiring Flexibility and Developing Rich Understanding of Integer Addition and Subtraction

We believe that acquiring flexibility when adding and subtracting integers should be a goal for every student. We found that flexibility is not only theoretically but also practically important: The more flexible students were in using the WoRs, the more accurate they were. Although we found a moderate-to-strong correlation between flexibility and accuracy, we did not establish causality. However, in other content areas, we know that students can learn to become more flexible (Berk et al., 2009; Star & Rittle-Johnson, 2008; Star & Seifert, 2006), and we suspect that the same may be true for integer addition and subtraction. In this article, we provide an operational characterization of flexibility that may serve as a guide for teachers to provide opportunities for their students to become more flexible in their approaches to solving integer addition and subtraction problems, thus providing opportunities for students to recognize features of problems that might make one approach more efficient than another.

To help develop flexibility, we suggest that more time be devoted to teaching about signed numbers. Given that learning about signed numbers has the potential to provide opportunities for students to explore and learn about new number systems; engage with structure and equivalent expressions; support students in learning algebra; and justify, reason, and conjecture, we suggest that additional time spent learning to grapple with integers will pay dividends in other content areas as well. We do not suggest a particular length of time for the study of integers but maintain that the current time devoted to the topic is insufficient.

A Developmental Trajectory of Learning About Integers

At the outset of our work, we wondered whether we would identify a hierarchical progression of relatively sophisticated WoRs in which one way of reasoning would replace another over time. Our cross-sectional findings indicate, however, that one productive way of reasoning is not necessarily replaced by others. Instead, powerful integer reasoning entails being able to flexibly invoke each of the four productive WoRs. Although CT students used a computational WoR more often than the other WoRs, they also productively used the other three productive WoRs, and they tended to use particular WoRs flexibly by attending to specific features of the open number sentences. That said, we view the role of the fifth WoR, emergent, as critical for teachers and researchers to understand and build from to facilitate students' sense-making approaches while they develop understanding of the other four WoRs.

We return now to Sam's and Ann's strategies provided at the beginning of the article to share how emergent reasoning might be leveraged to promote more sophisticated WoRs. Recall that Sam, a second grader, claimed that $6 + \square = 4$ had no solution because addition should make the sum the same or larger. Ann, a seventh grader, also recognized that addition made larger, but her understanding was more nuanced. Ann shared that her answer was -2 "because . . . you can't have a positive number [points to the blank] to get a number that is less than the first number. So you would have to have a negative number right there [points to the blank]." The difference between Sam's emergent WoR and Ann's formal WoR is not that Ann's reasoning drew on a replacement of, or on a conceptualization entirely different from, Sam's reasoning. On the contrary, Sam's claim formed the basis for Ann's response when Ann argued, in essence, that addition makes larger.

One important difference between Sam's reasoning and Ann's is that Ann bounded her claim regarding *the conditions under which* addition makes larger when she shared that "you can't have a *positive* [emphasis added] number [points to the blank] to get a number that is less than the first number." She extended this reasoning further by sharing her knowledge of the conditions under which addition can make smaller: "So you would have to have a negative number right there [points to the blank]." Sam's response reflects reasoning that we heard from almost every second and fourth grader we interviewed. Moreover, kernels of Sam's response were present in almost every instance of the strategy within the formal WoR, *inferring the sign*. Rather than attempting to replace Sam's reasoning about addition and subtraction, we recommend treating this emergent milestone as one that forms an important basis for the development of a more sophisticated response.

Limitations and Final Thoughts

Although we were able to answer some questions about students' approaches to solving integer addition and subtraction problems, others emerged. For example, we wondered whether and how success rates and approaches would change when PIN and CT students solved problems with magnitudes larger than 20 or when noninteger real numbers were used. Additionally, we investigated the approaches of college-track students in Grade 11, but we wonder how the results would have differed if we had included a representative sample of all 11th-grade students in our study. In future work, we see a need to explore students' approaches for integer multiplication and division problems and on integer algebraic expressions (e.g., $-4(x - 5)$).

In closing, we used the WoRs Framework to share how a broad cross-section of students solved addition and subtraction open number sentences; we identified three problem types that have the potential to evoke WoRs; and we conducted cross-sectional analyses to document the importance of flexibility for students' understanding of signed numbers. On the bases of our findings, we believe that no single best model or way of reasoning leads to students' success in solving open number sentences. Instead, we encourage teachers and researchers to come to understand the WoRs and the problem types to support students in developing productive WoRs about integers addition and subtraction problems, recognizing that even the emergent WoR has seeds from which productive WoRs can grow.

References

Behrend, J. L., & Mohs, L. C. (2005). From simple questions to powerful connections: A two-year conversation about negative numbers. *Teaching Children Mathematics*, *12*(5), 260–264.

Berk, D., Taber, S. B., Gorowara, C. C., & Poetzl, C. (2009). Developing prospective elementary teachers' flexibility in the domain of proportional reasoning. *Mathematical Thinking and Learning*, *11*(3), 113–135. doi:10.1080/10986060903022714

Bishop, J. P., Lamb, L. L. C., Philipp, R. A., Schappelle, B. P., & Whitacre, I. (2011). First graders outwit a famous mathematician. *Teaching Children Mathematics*, *17*(6), 350–358.

Bishop, J. P., Lamb, L. L., Philipp, R. A., Whitacre, I., & Schappelle, B. P. (2014). Using order to reason about negative numbers: The case of Violet. *Educational Studies in Mathematics*, *86*(1), 39–59. doi:10.1007/s10649-013-9519-x

Bishop, J. P., Lamb, L. L., Philipp, R. A., Whitacre, I., & Schappelle, B. P. (2016a). Leveraging structure: Logical necessity in the context of integer arithmetic. *Mathematical Thinking and Learning*, *18*(3), 209–232. doi:10.1080/10986065.2016.1183091

Bishop, J. P., Lamb, L. L., Philipp, R. A., Whitacre, I., & Schappelle, B. P. (2016b). Unlocking the structure of positive and negative numbers. *Mathematics Teaching in the Middle School*, *22*(2), 84–91. doi:10.5951/mathteacmiddscho.22.2.0084

Bishop, J. P., Lamb, L. L., Philipp, R. A., Whitacre, I., & Schappelle, B. P. (in press). Students' thinking about integer open number sentences. In L. Bofferding & N. M. Wessman-Enzinger (Eds.), *Exploring the integer addition and subtraction landscape: Perspectives on integer thinking*. Cham, Switzerland: Springer.

Bishop, J. P., Lamb, L. L., Philipp, R. A., Whitacre, I., Schappelle, B. P., & Lewis, M. L. (2014). Obstacles and affordances for integer reasoning: An analysis of children's thinking and the history of mathematics. *Journal for Research in Mathematics Education*, *45*(1), 19–61. doi:10.5951/jresematheduc.45.1.0019

Blöte, A. W., Van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology*, *93*(3), 627–638. doi:10.1037/0022-0663.93.3.627

Bobis, J., Clarke, B., Clarke, D., Thomas, G., Wright, R., Young-Loveridge, J., & Gould, P. (2005). Supporting teachers in the development of young children's mathematical thinking: Three large scale cases. *Mathematics Education Research Journal*, *16*(3), 27–57. doi:10.1007/BF03217400

Bofferding, L. (2010). Addition and subtraction with negatives: Acknowledging the multiple meanings of the minus sign. In P. Brosnan, D. Erchick, & L. Flevares (Eds.), *Proceedings of the 32nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. VI, pp. 703–710). Columbus, OH: The Ohio State University.

Bofferding, L. (2014). Negative integer understanding: Characterizing first graders' mental models. *Journal for Research in Mathematics Education*, *45*(2), 194–245. doi:10.5951/jresematheduc.45.2.0194

Bofferding, L., & Richardson, S. E. (2013). Investigating integer addition and subtraction: A task analysis. In M. Martinez & A. Castro Superfine (Eds.), *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 111–118). Chicago, IL: University of Illinois at Chicago.

Carney, M. B., Smith, E., Hughes, G. R., Brendefur, J. L., & Crawford, A. (2016). Influence of proportional number relationships on item accessibility and students' strategies. *Mathematics Education Research Journal*, *28*(4), 503–522. doi:10.1007/s13394-016-0177-z

Carpenter, T. P., Ansell, E., Franke, M. L., Fennema, E., & Weisbeck, L. (1993). Models of problem solving: A study of kindergarten children's problem-solving processes. *Journal for Research in Mathematics Education*, *24*(5), 428–441. doi:10.2307/749152

Carpenter T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (2014). *Children's mathematics: Cognitively Guided Instruction* (2nd ed.). Portsmouth, NH: Heinemann.

Carpenter, T. P., Fennema, E., Peterson, P. L., & Carey, D. A. (1988). Teachers' pedagogical content knowledge of students' problem solving in elementary arithmetic. *Journal for Research in Mathematics Education*, *19*(5), 385–401. doi:10.2307/749173

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, *26*(4), 499–531. doi:10.3102/00028312026004499

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school*. Portsmouth, NH: Heinemann.

Carpenter, T. P., & Moser, J. M. (1982). The development of addition and subtraction problem-solving skills. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 9–24). Hillsdale, NJ: Erlbaum.

Carpenter, T. P., & Moser, J. M. (1983). The acquisition of addition and subtraction concepts. In R. Lesh & M. Landau (Eds.), *The acquisition of mathematics concepts and processes* (pp. 7–44). New York, NY: Academic Press.

Case, R. (1996). Introduction: Reconceptualizing the nature of children's conceptual structures and their development in middle childhood. *Monographs of the Society for Research in Child Development*, *61*(1–2), 1–26. doi:10.1111/j.1540-5834.1996.tb00535.x

Chiu, M. M. (2001). Using metaphors to understand and solve arithmetic problems: Novices and experts working with negative numbers. *Mathematical Thinking and Learning*, *3*(2–3), 93–124. doi:10.1080/10986065.2001.9679970

Dowker, A. D., Flood, A., Griffiths, H., Harriss, L., & Hook, L. (1996). Estimation strategies of four groups. *Mathematical Cognition*, *2*(2), 113–135. doi:10.1080/135467996387499

Empson, S. B., & Levi, L. (2011). *Extending children's mathematics: Fractions and decimals.* Portsmouth, NH: Heinemann.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, *27*(4), 403–434. doi:10.2307/749875

Franke, M. L., Carpenter, T. P., Levi, L., & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal*, *38*(3), 653–689. doi:10.3102/00028312038003653

Fuson, K. C., Smith, S. T., & Lo Cicero, A. M. (1997). Supporting Latino first graders' ten-structured thinking in urban classrooms. *Journal for Research in Mathematics Education*, *28*(6), 738–766. doi:10.2307/749640

Gearhart, M., & Saxe, G. B. (2004). When teachers know what students know: Integrating mathematics assessment. *Theory Into Practice*, *43*(4), 304–313. doi:10.1207/s15430421tip4304_9

Hefendehl-Hebeker, L. (1991). Negative numbers: Obstacles in their evolution from intuitive to intellectual constructs. *For the Learning of Mathematics*, *11*(1), 26–32.

Henley, A. T. (1999). *The history of negative numbers* (Unpublished doctoral dissertation). South Bank University, London, United Kingdom.

Human, P., & Murray, H. (1987). Non-concrete approaches to integer arithmetic. In J. C. Bergeron, N. Herscovics, & C. Kieran (Eds.), *Proceedings of the Eleventh International Conference on the Psychology of Mathematics Education* (Vol. 2, pp. 437–443). Montreal, Canada: University of Montreal.

Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education*, *38*(3), 258–288.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press. doi:10.17226/9822

Lamb, L. L., Bishop, J. P., Philipp, R. A., Schappelle B. P., Whitacre, I., & Lewis, M. (2012). Developing symbol sense for the minus sign. *Mathematics Teaching in the Middle School*, *18*(1), 5–9. doi:10.5951/mathteacmiddscho.18.1.0005

Lamb, L. L., Bishop, J. P., Whitacre, I., Philipp, R.A., Schappelle, B. P., & Lewis, M. (2017). *Seventh-grade students' performance and ways of reasoning: Integer addition and subtraction problems* (White Paper). Retrieved from Center for Research in Mathematics and Science Education website: http://www.sci.sdsu.edu/CRMSE/projectz/documents/Integers%207th%20grade.pdf

LeFevre, J.-A., Bisanz, J., Daley, K. E., Buffone, L., Greenham, S. L., & Sadesky, G. S. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General*, *125*(3), 284–306. doi:10.1037/0096-3445.125.3.284

Linchevski, L., & Williams, J. (1999). Using intuition from everyday life in "filling" the gap in children's extension of their number concept to include the negative numbers. *Educational Studies in Mathematics*, *39*(1–3), 131–147. doi:10.1023/A:1003726317920

Moses, R. P., Kamii, M., Swap, S. M., & Howard, J. (1989). The algebra project: Organizing in the spirit of Ella. *Harvard Educational Review*, *59*(4), 423–443. doi:10.17763/haer.59.4.27402485mqv20582

Murray, J. C. (1985). Children's informal conceptions of integer arithmetic. In L. Streefland (Ed.), *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education* (Vol. 1, pp. 147–153). Utrecht, the Netherlands: State University of Utrecht.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

Patton, M. Q. (2015). Qualitative research & evaluation methods: Integrating theory and practice (4th ed.). Thousand Oaks, CA: Sage.

Peled, I. (1991). Levels of knowledge about signed numbers: Effects of age and ability. In F. Furinghetti (Ed.), *Proceedings of the Fifteenth Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 145–152). Assisi, Italy: International Group for the Psychology of Mathematics Education.

Peled, I., & Carraher, D. W. (2008). Signed numbers and algebraic thinking. In J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp. 303–328). New York, NY: Erlbaum.

Peled, I., Mukhopadhyay, S., & Resnick, L. B. (1989). Formal and informal sources of mental models for negative numbers. In G. Vergnaud, J. Rogalski, & M. Artigue (Eds.), *Proceedings of the 13th International Conference on the Psychology of Mathematics Education* (Vol. 3, pp. 106–110). Paris, France: International Group for the Psychology of Mathematics Education.

Sengupta-Irving, T., & Enyedy, N. (2015). Why engaging in mathematical practices may explain stronger outcomes in affect and engagement: Comparing student-driven with highly guided inquiry. *Journal of the Learning Sciences, 24*(4), 550–592. doi:10.1080/10508406.2014.928214

Smith, J. P., III. (1995). Competent reasoning with rational numbers. *Cognition and Instruction*, *13*(1), 3–50. doi:10.1207/s1532690xci1301_1

Sowder, J. (1992). Estimation and number sense. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 371–389). New York, NY: Macmillan.

Sowder, J. T. (2007). The mathematics education and development of teachers. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 157–223). Charlotte, NC: Information Age.

Star, J. R., Newton, K., Pollack, C., Kokka, K., Rittle-Johnson, B., & Durkin, K. (2015). Student, teacher, and instructional characteristics related to students' gains in flexibility. *Contemporary Educational Psychology*, *41*, 198–208. doi:10.1016/j.cedpsych.2015.03.001

Star, J. R., & Newton, K. J. (2009). The nature and development of experts' strategy flexibility for solving equations. *ZDM*, *41*(5), 557–567. doi:10.1007/s11858-009-0185-5

Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, *18*(6), 565–579. doi:10.1016/j.learninstruc.2007.09.018

Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology*, *31*(3), 280–300. doi:10.1016/j.cedpsych.2005.08.001

Steffe, L. P. (1992). Schemes of action and operation involving composite units. *Learning and Individual Differences*, *4*(3), 259–309. doi:10.1016/1041-6080(92)90005-Y

Steffe, L. P. (1994). Children's multiplying schemes. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 3–40). Albany: State University of New York Press.

Steffe, L. P. (2001). A new hypothesis concerning children's fractional knowledge. *Journal of Mathematical Behavior*, *20*(3), 267–307. doi:10.1016/S0732-3123(02)00075-5

Steffe, L. P., & Olive, J. (2010). *Children's fractional knowledge*. New York, NY: Springer. doi:10.1007/978-1-4419-0591-8

Stephan, M., & Akyuz, D. (2012). A proposed instructional theory for integer addition and subtraction. *Journal for Research in Mathematics Education*, *43*(4), 428–464. doi:10.5951/jresematheduc.43.4.0428

Taber, S. B. (2002). Go ask Alice about multiplication of fractions. In B. Litwiller (Ed.), *Making sense of fractions, ratios, and proportions* (pp. 61–71). Reston, VA: National Council of Teachers of Mathematics.

Thomaidis, Y., & Tzanakis, C. (2007). The notion of historical "parallelism" revisited: Historical evolution and students' conception of the order relation on the number line. *Educational Studies in Mathematics*, *66*(2), 165–183. doi:10.1007/s10649-006-9077-6

Vergnaud, G. (1982), A classification of cognitive tasks and operations of thought involved in addition and subtraction operations. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 39–59). Hillsdale, NJ: Erlbaum.

Vergnaud, G. (1983). Multiplicative structures. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 127–174). New York, NY: Academic Press.

Villaseñor, A., Jr., & Kepner, H. S., Jr. (1993). Arithmetic from a problem-solving perspective: An urban implementation. *Journal for Research in Mathematics Education*, *24*(1), 62–69. doi:10.2307/749386

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics*, *49*(3), 341–359. doi:10.1023/A:1020229023965

Vlassis, J. (2008). The role of mathematical symbols in the development of number conceptualization: The case of the minus sign. *Philosophical Psychology*, *21*(4), 555–570. doi:10.1080/09515080802285552

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*(4), 535–585. doi:10.1016/0010-0285(92)90018-W

Wessman-Enzinger, N. M. (2015). Developing and describing the use and learning of conceptual models for integer addition and subtraction of Grade 5 students. Available from ProQuest Dissertations and Theses database. (UMI No. 3725577)

Whitacre, I., Azuz, B., Lamb, L. L. C., Bishop, J. P., Schappelle, B. P., & Philipp, R. A. (2017). Integer comparisons across the grades: Students' justifications and ways of reasoning. *Journal of Mathematical Behavior*, *45*, 47–62. doi:10.1016/j.jmathb.2016.11.001

Whitacre, I., Bishop, J. P., Lamb, L. L. C., Philipp, R. A., Schappelle, B. P., & Lewis, M. L. (2012). Happy and sad thoughts: An exploration of children's integer reasoning. *Journal of Mathematical Behavior*, *31*(3), 356–365. doi:10.1016/j.jmathb.2012.03.001

Whitacre, I., Bishop, J. P., Philipp, R. A., Lamb, L. L., & Schappelle, B. P. (2014). Dollars and sense: Students' integer perspectives. *Mathematics Teaching in the Middle School*, *20*(2), 84–89. doi:10.5951/mathteacmiddscho.20.2.0084

Whitacre, I., Bouhjar, K., Bishop, J. P., Philipp, R. A., Schappelle, B. P., & Lamb, L. L. (2016). Regular numbers and mathematical worlds. *For the Learning of Mathematics*, *36*(2), 20–25.

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, *24*, 173–209. doi:10.2307/1167270

Authors

Lisa L. Lamb, School of Teacher Education and Center for Research in Mathematics and Science Education, San Diego State University, CRMSE, 6475 Alvarado Road, Suite 206, San Diego, CA 92120; lisa.lamb@sdsu.edu

Jessica Pierson Bishop, Department of Mathematics, Texas State University, 601 University Drive, San Marcos, TX 78666; jbishop01@txstate.edu

Randolph A. Philipp, School of Teacher Education and Center for Research in Mathematics and Science Education, San Diego State University, CRMSE, 6475 Alvarado Road, Suite 206, San Diego, CA 92120; rphilipp@sdsu.edu

Ian Whitacre, School of Teacher Education, Florida State University, 1114 West Call Street, G123 Stone Building, Tallahassee, FL 32306; iwhitacre@fsu.edu

Bonnie P. Schappelle, Center for Research in Mathematics and Science Education, San Diego State University, CRMSE, 6475 Alvarado Road, Suite 206, San Diego, CA 92120; bschappelle@sdsu.edu

Submitted October 5, 2016
Accepted February 7, 2017

APPENDIX
Integers Problem-Solving Interview

Introductory Questions

1) Name a big number. Can you name a bigger number?

2) Name a small number. Can you name a smaller number? If the child responds, “Zero,” ask, “Is there a number smaller than zero?”

3) Can you count backward, starting at 5? If child stops at 0 or 1, ask, “Can you keep counting back?”

4) What can you tell me about negative numbers? (Ask only if the student has previously mentioned the term *negative*.)

*Control Problems, Natural Numbers only Open Number Sentences**

5) 5 + 6 = □ 6) 4 + □ = 9 7) □ – 4 = 6 8) 8 – □ = 4

*Open Number Sentences***

(The CT, PIN, and BIN participant groups solved each of the following 25 problems. The NEN students were posed only questions 9–14 and 23–28 to reduce their time and, potentially, stress when solving problems with negative numbers for which they had no background.)

9) 3 – 5 = □	10) 6 + □ = 4	11) 5 – □ = 8
12) □ + 6 = 2	13) -3 + 6 = □	14) -8 – 3 = □
16) -2 + □ = 4	17) □ – 5 = -1	18) -9 + □ = -4
19) -2 – □ = -8	20) -5 + □ = -8	21) -3 – □ = 2
22) -8 – □ = -2	23) -8 + □ = 0	24) -5 + -1 = □
25) -5 – -3 = □	26) 6 – -2 = □	27) 6 + -3 = □
28) 3 + □ = 0	30) -5 – -5 = □	31) -7 – -9 = □
32) □ + -7 = -3	33) □ + -2 = -10	34) 3 – □ = -6
35) -2 – 7 = □		

*Findings presented in the article did not include percentage correct for the control problems. Students solved the control problems 5–8 so that we could more reliably attribute any challenges in solving problems 9–35 to difficulties with negative numbers rather than to challenges in understanding the structure of open number sentences. CT students correctly solved 100% of the control problems; PIN students, 98%; BIN students, 95%; and NEN students, 88%.

**Questions 15 and 29 are not open number sentences.