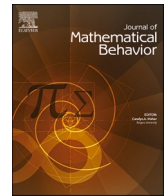




Contents lists available at ScienceDirect

Journal of Mathematical Behavior

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

Flexibility across and flexibility within: The domain of integer addition and subtraction

Lisa Lamb^{a,*}, Jessica Bishop^b, Ian Whitacre^c, Randolph Philipp^a^a San Diego State University, 6475 Alvarado Road, Suite 206, San Diego, CA 92120, USA^b Texas State University, Department of Mathematics, Texas State University, 601 University Drive, San Marcos, TX 78666, USA^c Florida State University, 1114 W. Call St., G123 Stone Bldg., Tallahassee, FL 32306-4459, USA

ARTICLE INFO

Keywords:

Number concepts and operations

Cognition

Flexibility

Adaptive expertise

Strategy variability

Integers

ABSTRACT

To better understand the role that flexibility plays in students' success on integer addition and subtraction problems, we examined students' flexibility when solving open number sentences. We define flexibility as the degree to which a learner uses more than one strategy to solve a single task when prompted, as well as the degree to which a learner changes strategies when solving a range of tasks to accommodate task differences. We introduce the categorizations of flexibility within and flexibility across to distinguish these two ways of operationalizing flexibility. We examined flexibility and performance within and among three groups of students — 2nd and 4th graders who had negative numbers in their numerical domains, 7th graders, and college-track 11th graders. Profiles of five students are shared to provide insight in relation to the quantitative findings.

1. Introduction

Since the 1980s, researchers have found that children and adults exhibit adaptive strategy choices that reflect implicit attention to problem difficulty, numbers in the problem, the child's competence with a strategy, and more (e.g., Lemaire & Siegler, 1995; Siegler, 1987; Siegler & Lemaire, 1997; Smith, 1995; Sowder, 1992; Verschaffel et al., 2009). Further, in the area of single-digit whole-number addition, subtraction, and multiplication, a child will continue to use a variety of strategies to solve problems over a period of years, decreasing the use of more time-intensive, less successful strategies and increasing the use of more efficient strategies to solve a given task (Lemaire & Siegler, 1995; Siegler et al., 1996).

We wondered whether a similar trajectory of strategy use would hold in the domain of additive integer arithmetic, and so in this article, we share our cross-sectional analysis of students' flexibility in the content domain of additive integer arithmetic. For us, flexibility captures the idea that students may be able to use multiple methods when solving integer addition and subtraction problems. More specifically, we define flexibility as the degree to which a learner uses more than one strategy to solve a single task when prompted, as well as the degree to which a learner changes strategies when solving a range of tasks to accommodate task differences. To begin, we share our conceptual framework of ways of reasoning and problem types for integer addition and subtraction. Next, we explore previous research on flexibility and provide two subtypes of flexibility, *flexibility across* and *flexibility within*, which in some studies are called flexibility, strategy variability, adaptive strategy use, adaptive expertise, and more.

* Corresponding author.

E-mail addresses: Lisa.Lamb@sdsu.edu (L. Lamb), jbishop01@txstate.edu (J. Bishop), iwhitacre@fsu.edu (I. Whitacre), Rphilipp@sdsu.edu (R. Philipp).

2. Conceptual framework: Ways of reasoning and problem types

In this section, we share a brief history with respect to the importance of investigating problem types, ways of reasoning, and their relationship. In [Section 2.1](#), we then share the range of approaches that students use when solving additive integer open number sentences, followed by a discussion in [Section 2.2](#) of problem-type categories for integer addition and subtraction.

Identifying ways of reasoning and problem types has a rich history in the field of mathematics education. For example, in seminal work, [Carpenter et al. \(1981\)](#) identified addition and subtraction problem types and strategies that young students use to solve story problems prior to formal instruction. The work was further articulated ([Carpenter & Moser, 1984](#)) and these basic research studies laid the foundation for applied research in classrooms, including the groundbreaking Cognitively Guided Instruction program of research (e.g., [Carpenter et al., 2014, 1989](#); [Fennema et al., 1996](#); [Franke et al., 2009](#)). The investigators of these basic research studies sought to understand students' informal mathematical ideas and used these ideas to document the relationship between students' strategies and problem structure. For example, [Carpenter et al. \(1981\)](#) found that when young children solved story problems involving natural numbers, problem structure influenced problem difficulty and strategies in ways that previous researchers had not articulated. For example, although previous researchers had identified missing addend problems as challenging for children who had already received formal instruction, [Carpenter and Moser \(1984\)](#) found that, prior to formal instruction, young children solved problems such as "Reid has 4 toy cars. His mom gives him more toy cars. Now he has 11 toy cars. How many cars did Reid's mom give to Reid?" by following the action in the story, and approaching this problem as a joining problem with the change value unknown (join change unknown, [Carpenter et al., 2014](#)). In contrast, in previous research, researchers expected students to transform this problem into a subtraction problem and solve 11–4.

Incorporating students' strategies to help identify different additive problem structures was a critical decision point in developing what are now widely accepted frameworks for problem types. Further, the approach of identifying both students' approaches to solving problems and the distinct ways that problem types influenced those strategies is now well documented in many areas beyond natural-number problem solving, including fractions (e.g., [Empson and Levi, 2011](#)), relational thinking (e.g., [Carpenter et al., 2005](#)), and shape identification (e.g., [Clements et al., 1999](#)). We adopted a similar approach in our work by identifying ways of reasoning and problem types in the domain of integer addition and subtraction and the relationship between the two.

2.1. Ways of reasoning (WoRs)

In previous work, we documented five broad Ways of Reasoning (WoRs) that students in elementary, middle, and high school use when solving integer open number sentences ([Bishop et al., 2014, 2018](#); [Lamb et al., 2018](#)). The WoRs are order-based, analogy-based, formal, computational, and emergent and are characterized in [Table 1](#). Other scholars have also documented one or more of these WoRs ([Bofferding, 2014, 2019](#); [Hefendehl-Hebeker, 1991](#); [Human & Murray, 1987](#); [Murray, 1985](#); [Peled, 1991](#); [Stephan & Akyuz, 2012](#); [Wessman-Enzinger, 2015, 2019a](#)). For an example of an order-based WoR, to solve $-2 + 5$, a child might place a pen at -2 on a number line and move the pen five units to the right, landing at 3, taking advantage of the sequential order of numbers to solve the problem ([Bishop et al., 2014](#); [Stephen & Akyuz, 2012](#)). In contrast, a student might use an analogy-based WoR by comparing negative numbers to another entity. For example, to solve $-5 - 3$, a child may get -2 by subtracting $5 - 3$ to get 2, and sharing that, for this problem, negative numbers act like "regular" or positive numbers. Students who use a procedure or rule to solve a problem are using a

Table 1
Ways of reasoning framework (from [Lamb et al., 2018](#)).^a

Way of Reasoning		Definition
Evidence of Engagement With Negative Numbers	Order-based	One leverages the sequential and ordered nature of numbers to reason about a problem. Strategies include using the number line with motion as well as counting forward or backward by 1 s 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 bad guys). Analogy-based strategies are often tied to ideas about cardinality and understanding a number as having magnitude.
	Formal	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 by generalizing beyond a specific case to apply properties of classes of numbers, such as generalizations about zero.
	Computational	One uses a procedure, rule, or calculation to arrive at an answer. 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 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.

^a For additional details, see [Lamb et al. \(2018\)](#). See also an early version of this framework ([Bishop et al., 2014](#)).

computational WoR. For example, a student is using a computational WoR when invoking a rule, such as changing $6 - -2$ to $6 + +2$ using what many students call the *Keep-Change-Change* (KCC) rule, such that one **keeps** the sign of the minuend (or first addend), **changes** the operation, and **changes** the sign of the subtrahend (or second addend in the case of addition). Students used a formal WoR when invoking properties of number, sometimes to solve a problem that they were previously unable to solve or explain (Bishop et al., 2016). For example, students might compare a target problem for which the answer is not known (e.g., $6 + -2 = _$) to another similar problem for which the answer is known (e.g., $-2 + 6 = 4$) and reason by comparing similarities or differences in the problems. In this example, students would share that the order of the addends does not matter and so the answer to both problems must be 4 because of the commutative property of addition. Formal WoRs differ from computational WoRs in that formal.

WoRs provide valid mathematical justification that leverages mathematical structure and fundamental properties. For example, many students solved the problem $6 - 2$ using only a Computational WoR as described above. However, some students used a Formal WoR by comparing $6 - -2$ to $6 - 2$ and using the inverse relationship between 2 and -2. One student argued that subtracting the opposite of two should have the opposite result as subtracting a “regular” 2 and since the answer could not be 4, it must be 8. Emergent WoRs reflect attempts to solve additive integer problems within the domain of whole numbers. For example, a child might solve $-2 + 5$ and get 7, ignoring the negative sign. Or, 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 (e.g., $6 + _ = 4$) has no answer. Interestingly, early mathematicians made the same kinds of claims (Hefendehl-Hebeker, 1991).

The first four WoRs, order-based, analogy-based, formal, and computational, are all mathematically valid. That is, using these WoRs may not always produce a correct answer, but each WoR has an accepted mathematical basis. The fifth WoR, *emergent*, while developmentally important (see Lamb et al., 2018), includes only strategies that are restricted locally to the domain of whole numbers. Because emergent approaches, although sensible and understandable, are not mathematically valid, we did not include them in our flexibility calculations.

2.2. Problem types

Researchers have grouped additive integer problems in a variety of ways (Aqazade & Bofferding, 2021; Bishop et al., 2022; Bofferding & Aqazade, 2018; Bofferding & Wessman-Enzinger, 2017; Glancy & Pettis, 2017; Lamb et al., 2018; Wessman-Enzinger & Tobias, 2022). In previous research (Lamb et al., 2018) we analyzed 975 responses to 25 open number sentences and identified four problem types for integer addition and subtraction open number sentences. The problem types were grouped on the bases of their relative difficulty, the WoRs they tended to evoke, and their mathematical structure. The four problem types are change-negative all-negative, change-negative not all-negative, change-positive cross-zero, and change-positive negative-side (see Table 2).¹ We identified change-negative not-all-negative problems and change-positive negative-side problems as especially challenging (Lamb et al., 2018). Others have also identified problem types and many of our categories overlap. For example, Glancy and Pettis (2017) identified ten integer problem types and found that three types (positive minus negative equals positive and negative minus negative equals positive, a subset of what we call change-negative not all-negative; and negative minus positive equals negative, a subset of what we call change-positive negative-side) were statistically significantly more challenging for sixth-grade students than the other problem types. Bofferding and Wessman-Enzinger (2017) investigated students’ thinking about integer subtraction problems and found that negative-minus-negative-equals-negative problem types (a subset of what we call change-negative all-negative) were easier to solve than other kinds of subtraction problems.

As described elsewhere (Lamb et al., 2018), prior to school-based instruction, these problem types typically evoke particular WoRs: Change-negative-all-negative problems tend to evoke analogy-based reasoning; change-positive problems tend to evoke order-based reasoning; and change-negative not-all-negative problems tend to evoke emergent reasoning (Lamb et al., 2018). For example, before school-based instruction, whereas 87 % of the students who knew something about negative numbers answered the change-negative-all-negative problem $-5 - -3$ correctly, only 26 % answered the change-negative-not-all-negative problem $-7 - -9$ correctly. Further, 54 % of these students answered the change-positive-cross-zero problem $-2 + _ = 4$ correctly, but only 21 % answered the change-positive-negative-side problem $-9 + _ = -4$ correctly. As shared in Section 2, when problem type categories are informed by students’ approaches, they can provide a foundation for building from and extending students’ mathematical ideas. By asking participants in our study to solve problems from every problem-type category, we were able to examine participants’ degrees of flexibility across problems, precisely because the problem types have the potential to evoke differing ways of reasoning.

3. Flexibility

In this section, we begin by briefly sharing research related to flexibility, and then introduce two terms to support the field’s understanding of the different ways that flexibility has been operationalized: *flexibility within* and *flexibility across*. We follow these subsections with a discussion of research about the relationship between flexibility and performance, and end this section by sharing what is known about flexibility in the domain of integer addition and subtraction.

For decades researchers have investigated students’ and experts’ flexibility in many content areas. In the 1990s, several researchers found that experts or students who were competent reasoners demonstrated flexibility with estimation (e.g., Dowker et al., 1996;

¹ Although the problem types in Table 2 are the same as reported in Lamb et al. (2018), the names have been updated. Specifically, what was called “counterintuitive” is now called, “change-negative not-all-negative” to reflect the mathematical features of the problem.

Table 2
Open Number Sentences Categorized by Problem Type.

Change Negative		Change Positive	
All Negatives	Not All negatives	Cross Zero	Negative Side
All terms are negative	2nd term is negative and at least one of the other terms is positive (or 0 for addition).	2nd term is positive. The start and result have opposite signs (or one of them is 0).	2nd term is positive. The start and result are both negative.
$-5 + -1 = _$	$6 + -3 = _$	$3 - 5 = _$	$-8 - 3 = _$
$-5 - -3 = _$	$-7 - -9 = _$	$-3 + 6 = _$	$-9 + _ = -4$
$_ + -2 = -10$	$_ + -7 = -3$	$-2 + _ = 4$	$-2 - 7 = _$
$-8 - _ = -2$	$6 - -2 = _$	$_ - 5 = -1$	$-2 - _ = -8$

Sowder, 1992), single-digit multiplication (Le Fevre et al., 1996; Lemaire & Siegler, 1995), and reasoning with rational numbers (Smith, 1995). For example, Smith (1995) found that competent reasoning with fractions included both general strategies typically taught in school (e.g., finding a common denominator to compare fractions) and specific strategies that applied only "in numerically specific situations" (p. 38). Since that time, others have examined students' and experts' flexibility² with respect to solving linear equations (e.g., Newton et al., 2010; Rittle-Johnson & Star, 2007; Star & Newton, 2009), proportional reasoning (e.g., Berk et al., 2009), representations (e.g., Nistal et al., 2012, 2014), and arithmetic (e.g., Heirdsfeld & Cooper, 2002; Shaw et al., 2020; Siegler, 1987). One of the striking features of this body of research is the number of content areas in which flexibility has been studied. We suggest that this diversity in content and topics shows that flexibility is a critical component of mathematical reasoning across the curriculum.

3.1. Flexibility-within and flexibility-across problems

Scholars agree that flexibility involves having and using multiple methods for solving problems, and this idea has been operationalized in various ways. With one approach, researchers considered the degree to which participants can solve a single problem in more than one way, an approach that we call *flexibility within* a problem. With another approach, researchers considered the degree to which participants change from one strategy to another across a set of problems, an approach that we call *flexibility across*.³ Some researchers have investigated both approaches (e.g., Berk et al., 2009; Blöte et al., 2001). In this article, we characterize flexibility as comprising both flexibility within and flexibility across. In the following sections, we share other researchers' characterizations and ground ours in this literature.

3.1.1. Flexibility Within

Many researchers have examined flexibility within a given problem (e.g., Newton et al., 2010; Rittle-Johnson & Star, 2007; Shaw et al., 2020; Xu et al., 2017). For example, Shaw et al. (2020) had college students solve the same problem, $820 - 410$, repeatedly for eight minutes, asking for a different strategy on each attempt. Newton et al. (2010) and Rittle-Johnson and Star (2007) asked students to solve the same algebra problem in more than one way, as well as to identify whether a nonconventional approach was a "very good way," "OK to do, but not a very good way," or "not OK to do" (Rittle-Johnson & Star, p. 565). Examining whether students can solve a problem in more than one way aids researchers who investigate the mathematical ideas of students who are likely to have been taught standard (also referred to as conventional or canonical) approaches and who are then likely to primarily use those standard approaches. By asking for multiple ways to solve the same problem, researchers can assess the degree to which students are able to generate and use nonstandard (what other researchers also call nonconventional or innovative) approaches (Hästö et al., 2019; Star et al., 2022; Xu et al., 2017).

3.1.2. Flexibility across

Other researchers have examined flexibility across problems (e.g., Dowker et al., 1996; Heirdsfeld & Cooper, 2002; Lemaire & Siegler, 1995; Nistal et al., 2014; Siegler, 1987). For example, Dowker et al. (1996) interviewed participants belonging to one of four groups: mathematicians, professional accountants, undergraduate English majors, and undergraduate psychology majors. They assessed "within-individual" variability by counting the number of strategies each participant used across a set of estimation tasks. Lemaire and Siegler (1995) interviewed 20s-grade students about single-digit multiplication at three timepoints in a school year. They found that, on average, each child used a little more than three strategies per session. Similarly, Siegler (1987) found that when 5- and 6-year old children responded to 25 Separate Result Unknown problems (Carpenter et al., 2014), three fourths of them used 3 or 4

² Many researchers referred to procedural flexibility, and others referred to representational flexibility. We do not identify flexibility for additive integer arithmetic as "procedural" because solving additive integer arithmetic problems requires many kinds of understanding. Restricting our findings to the category of procedural would not fully account for the understanding the students displayed.

³ Other researchers have called these strategy changes *adaptive strategy use* (e.g., Siegler, 1987). We use two new terms, flexibility across and flexibility within, to reflect these distinctive kinds of flexibility.

strategies with varying frequency across the problems. Thus, these researchers assessed flexibility across⁴ and found that participants changed strategies within the same problem-solving interview. Relatedly, Siegler (2016) has found that this strategy variability persists over time rather than getting replaced by new strategies. He referred to this way of thinking about children's ideas over time as an *overlapping waves theory* and wrote, "The overlapping waves model posits that on most problems with which children have experience, several ways of thinking and acting coexist and compete over prolonged periods of time" (p. 130).

Examining flexibility across aids researchers who want to understand the degree to which learner characteristics and task characteristics influence a change in strategy use. We now turn to a discussion of what researchers have learned about the relationship between flexibility and performance.

3.2. Performance and flexibility

Researchers have examined students' performance in relation to their degree of flexibility (Berk et al., 2009; Nistal et al., 2012; Rittle-Johnson & Star, 2007). These studies sometimes included an intervention to support the development of increased flexibility. For example, Rittle-Johnson and Star (2007) examined 7th-grade students' learning under two conditions. In the comparison condition, the 7th-grade students compared worked solution methods side by side and then were prompted to consider why the learner might choose one way instead of the other. In the sequential condition, learners read the same solution methods as in the comparison condition, but they viewed them sequentially. After reviewing each solution, they were prompted to respond to 1–2 questions. The comparison condition supported both improved performance of procedural knowledge and improved flexibility within for learning how to solve equations. With respect to proportional reasoning, Berk et al. (2009) designed a comparing-strategies intervention into a methods course to improve prospective teachers' flexibility, both across and within. They found that prospective teachers exhibited limited flexibility prior to participating in the methods course but that they learned to become more flexible and more accurate, even six months after the intervention. On the basis of these studies, the field has evidence that greater flexibility is positively related to knowledge and accuracy.

3.3. Flexibility and the domain of additive integer arithmetic

Additive integer arithmetic has the potential to provide another rich domain for exploring students' flexibility for three reasons. First, integer computation, especially subtraction, is known to be especially challenging for middle and high school students (Bofferding & Richardson, 2013; Gallardo, 2002; Lamb et al., 2018; Nurnberger-Haag, 2018; Vlassis, 2002), and a lack of proficiency with integer addition and subtraction is related to students' lack of success in algebra (Moses et al., 1989; Thomaidis & Tzanakis, 2007; Vlassis, 2002). Thus, gaining better understanding of the role that flexibility plays in this domain provides the field opportunities to make more informed decisions about productive learning opportunities for students. Second, researchers have documented both problem types and Ways of Reasoning (WoRs) that students use to solve additive integer problems (Bishop et al., 2022; Bofferding & Wessman-Enzinger, 2017; Glancy & Pettis, 2017; Lamb et al., 2018). In particular, several researchers have documented a *variety* of ways that students reason, and these studies serve as important precursors to this current study on flexibility (e.g., Bishop et al., 2018; Bofferding et al., 2018; Bofferding & Wessman-Enzinger, 2017; Wessman-Enzinger, 2015, 2018). Awareness of these WoRs enables researchers to examine students' use of a range of WoRs across different types of additive integer problems. Third, flexibility across is especially salient in the domain of additive integer arithmetic. For example, when a child solves $-2 + 5$ by counting-on five units from -2 to get 3, and then solves $-8 + -1$ by recognizing that 8 negative 1 s plus 1 negative 1 will result in an answer of 9 negative 1 s, the child is using different WoRs that are compatible with different problem types. Within the domain of integer arithmetic, the characteristic of *flexibility across* problems was informally recognized decades ago when 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 -4×5 by extrapolating from known number facts. (p. 149)

Recently we documented statistically significant relationships between problem types and the WoRs used to solve problems of each type (Lamb et al., 2018). As a next step, understanding (a) the degree to which students demonstrate flexibility within and flexibility across open number sentences, (b) how and to what degree the prevalence of flexibility changes between participant groups, and (c) the relationship between flexibility and accuracy within participant groups is important. In the Methods section, we describe how we operationalize flexibility, and how that operationalization differs in important ways from that used in our previous work (Lamb et al., 2018).⁵

4. Research questions

In this investigation, we advance the field's understanding of additive integer reasoning by explicitly examining flexibility within

⁴ Our term. The researchers of these studies used different terms to describe this phenomena: within-individual variability and adaptive strategy use.

⁵ In previous work, we operationalized flexibility across differently than we do herein (Lamb et al., 2018). Further, the 2018 article did not include flexibility-within findings or student profiles.

and flexibility across in this domain, both prior to and after school-based instruction. Through this basic research, we can gain a better understanding of flexibility in the domain of additive integer arithmetic and thus add to the research base of both students' understanding of integers and the construct of flexibility more generally (Herbst et al., 2022). Thus, we investigated the following three related research questions:

1. To what degree do students demonstrate flexibility across and flexibility within when solving additive integer arithmetic problems, and to what degree are these two components related?
2. How and to what degree does the prevalence of flexibility change between and among elementary, middle, and high school students?
3. To what degree are flexibility and performance related within additive integer problems?

5. Methods

This study is part of a larger project in which our goal was to understand K–12 students' conceptions of integers and integer arithmetic. We used an exploratory mixed methods research design and conducted and analyzed 90 pilot interviews to develop our problem-solving interview and initial WoR frameworks. After finalizing the interview tasks and protocol, we conducted 160 additional interviews and analyzed them both qualitatively and quantitatively (Creswell & Plano Clark, 2017).

5.1. Background and participants

Data for this study included clinical interviews with 160 students from 11 schools (3 elementary, 3 middle, 1 K–8, and 4 high schools) in the Western United States. The schools were selected to represent socioeconomic diversity. The students participated in individual interviews focused on additive integer arithmetic. We conducted clinical interviews with students in Grades 2, 4, 7, and 11 (40 children from each grade level). We chose these grade levels for the following reasons: Students in Grades 2 and 4 had yet to receive school-based integer instruction; students in Grade 7 had completed integer instruction; and students in Grade 11 were enrolled in a precalculus or calculus course and, because of their course taking, were deemed to be college-track high school mathematics students. We named the participant 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) ($n = 40$, 7th-grade students who had recently completed instruction in integers), Before Instruction with Negatives students (BINs) ($n = 39$, 2nd and 4th graders with negative numbers in their numerical domains), and No Evidence of Negatives (NENs) ($n = 41$, 2nd and 4th graders who provided no evidence of having negative numbers in their numerical domains). Group placements for 2nd and 4th graders were made on the basis of responses to an initial set of three tasks in the interview.⁶ The BIN group included 13 Grade 2 students and 26 Grade 4 students who, on the basis of responses to the three 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 the three tasks, provided no evidence of having knowledge of negative numbers.⁷ Because our focus is on understanding the flexibility of students who had some familiarity with negative numbers, we restrict our findings in this paper to the CT, PIN, and BIN groups.

5.2. Clinical Interview

The videotaped 45–90-minute clinical problem-solving interviews (Ginsburg, 1997) were conducted at the students' school sites. Although we sought to understand and follow the child's thinking during the interviews, the interviews were standardized. The interview had four task categories: introductory questions, open number sentences, contextualized problems, and comparison problems. Findings shared in this manuscript are from responses to the 25 open number sentences we posed (see examples in Table 2) to the BIN, PIN, and CT students. The interviewer provided unifix cubes and number lines to the students and shared that the students could solve the problem in any way they wanted, including using the cubes, number lines, fingers, or no tool. The problems were presented one at a time, and each student was asked to read the problem aloud before sharing an answer or explanation. In the introduction to the interview, the interviewer told the student that every time an answer was provided, the interviewer would ask, "How did you think about that?" If needed, interviewers posed follow-up questions to clarify students' approaches, such as, "I'm not sure I know what you just did. Can you tell me more about how you thought about that?" These questions were meant to uncover how the student was thinking and not to encourage any particular solution or WoR. Every 20–30 min, students were asked whether they would like a stretch or restroom break.

5.3. Coding and analysis

The interviews were coded at the problem level for both correctness and the underlying WoR the child used. The interview items

⁶ Questions included (a) Name a small number. Can you name a smaller number? (b) Can you count back starting at 5? (c) Solve $3 - 5 = _$. These questions had specific follow-up prompts to identify whether students had encountered negative numbers.

⁷ For a detailed analysis of 2nd graders' conceptions prior to having negative numbers in their numerical domains, see Bofferding et al. (2018).

and the coding frameworks for the WoRs were developed through analyses of 90 pilot interviews.⁸ Each response to the 25 open number sentences was assigned a WoR code and a subcode. Of the 160 interviews, 42 (or 26.25 %) were double coded, and interrater agreement was 92 % at the WoR level. Sometimes students used more than one WoR when solving a problem. For example, on the problem $-8 - 3$, one student first invoked a rule to rewrite the problem as $-8 + -3$, and then used an analogy of comparing negative numbers to positive numbers to get the answer of -11 . Because the student used both computational and analogy-based reasoning together to obtain a solution, the WoRs coded for this problem were computational and analogy-based and were coded as *Bundled*. Despite the use of two WoRs, we considered these kinds of responses as one solution because of the way students integrated and combined the WoRs to obtain an answer.

5.3.1. Measuring flexibility

As shared in Section 4, we assessed (a) the degree to which participants exhibited flexibility across problems and (b) the degree to which participants exhibited flexibility within problems.

5.3.1.1. Flexibility across problems. To operationalize flexibility across problems, we first computed the frequency of use for the following WoRs for each student: analogy-based, order-based, formal, and computational. We then operationalized flexibility across by determining the number of WoRs each student used at least three times across the 25 open number sentences. Newton et al. (2010) found that students may have knowledge of approaches but use them infrequently (see also Blöte et al., 2001; Star & Rittle-Johnson, 2008); consequently, we set a minimum number of uses for each WoR to increase the likelihood that we were capturing WoRs that the students both knew and used. This characterization enabled us to create a student's flexibility-across score by including those WoRs a student used routinely, without requiring prodigious use of that WoR. After computing a flexibility-across score for each student, to answer Research Questions 2 and 3, we used correlations (across all students and within each participant group) to explore two relationships: between flexibility across and flexibility within and between flexibility and performance on the 25 open number sentences. We further explored research questions 2 and 3 by sharing profiles of five students. We selected students to profile whose flexibility and performance provided a range in scores and offered texture to the quantitative results. We selected 1 BIN student, 3 PIN students, and 1 CT student to profile. The BIN and CT students were reasonably representative of the students in each of these participant groups. The 3 PIN students represent a range within the PIN group. We share additional profiles in this group because the PIN students had the greatest variability in their flexibility and accuracy and these profiles illustrate that variation.

5.3.1.1.1. Previous operationalization of flexibility. We previously reported on students' flexibility across when solving additive integer open number sentences (Lamb et al., 2018). In that research, we operationalized flexibility across by measuring what might be thought of as how strategic students were rather than how flexible they were. In the previous study, we adopted Star and Newton (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 given problem and a particular problem-solving goal" (p. 558). In the 2018 study, we interpreted "selectively choos[ing] the most appropriate [strategies] for a given problem" by assigning credit to a student who used a particular Way of Reasoning only if that use aligned with the problem type likely to elicit that WoR. For example, in our prior operationalization of flexibility, a student was determined to have used an order-based way of reasoning only if the student used that WoR on at least three change-positive problems. That is, if a student used order-based reasoning on, say, two change-positive problems and one change-negative problem, the student would *not* have been determined to have used order-based reasoning as a way of reasoning, because the use of order-based reasoning on the change-negative problem was not considered "appropriate" from a strategic perspective. This characterization of flexibility led us to disregard many of the students' uses of Ways of Reasoning that made sense to them. In the intervening years, we have come to further recognize the value of students' mathematical ideas especially in relation to assessing their flexibility (cf. Nistal et al., 2012). In particular, students may have a preference for or be especially competent with a particular way of reasoning, and we want to account for that way of reasoning in our current investigation. In our previous operationalization of flexibility, we did not account for these students' competencies. In our current operationalization (see 5.3.1.1), we do.

5.3.1.2. Flexibility within. To assess flexibility within, a total of 46 PIN and CT students were explicitly prompted for a different way to solve two open number sentences, $-5 - -3$ and $-2 - 7$, after solving these problems using any approach they wanted. After solving $-5 - -3$, they were asked, "Do you have another way of thinking about what it means to subtract negative 3 from negative 5?" After solving $-2 - 7$, they were asked, "Sometimes there is another way to think about a problem. Could you find another way to solve this problem?" We called these responses *Prompted Secondary WoRs*. For each problem, we assigned a score of 1 if the students shared a secondary WoR and 0 if they did not. Thus, the range of possible scores was 0–2.

Interview lengths ranged from 45 to 90 min, and so interviewers judged whether to pose these additional prompts on the basis of the length of the interview and their perceptions about how long the child could continue the interview.⁹ To answer Research Questions 2 and 3, after computing a flexibility-within score for each student, we used correlations (across the group of students and within the PT and CT groups) to explore the relationships between flexibility across and flexibility within and between flexibility within and performance on the 25 open number sentences.

⁸ These 90 pilot interviews occurred prior to the 160 interviews we report on in this paper.

⁹ We take up this decision in the Limitations section.

5.3.1.3. Example of flexibility scoring. To provide an example of how scores were created for each participant, we provide an example of codes for Josh (a pseudonym). He used order-based reasoning on 14 problems, analogy-based reasoning on 1 problem, computational reasoning on 9 problems, and formal reasoning on 4 problems. Thus, his flexibility-across score was 3 (on a scale of 0–4), indicating his familiarity and recurrent use of order-based reasoning, computational reasoning, and formal reasoning. To assess flexibility within, Josh was prompted to provide a secondary WoR for both $-5 - -3$ and $2 - 7$. He was unable to provide a secondary WoR for $-5 - -3$ but was able to share a secondary WoR for $-2 - 7$. Thus, his flexibility-within score was 1 (on a scale from 0 to 2). He correctly answered 80 % of the open number sentences. See Table 3 for a list of measures gathered for participants.

6. Findings

In this section, we report quantitative findings about flexibility for additive integer arithmetic and share five profiles to situate and contextualize the quantitative findings. We begin by examining the number of different WoRs students in each participant group flexibly used across open number sentences. We then examine how the number of WoRs changed from BIN to PIN to CT group. We follow these analyses with analyses of Flexibility Within, the relationship between Flexibility Across and Flexibility Within, and the relationship between Flexibility and Performance. In Section 6.4, we showcase five profiles of students to provide insights into the quantitative findings related to each research question.

6.1. Flexibility across

To showcase students' flexibility-across problems, in Fig. 1, we share frequency percentages for flexibility-across scores of 0, 1, 2, 3, and 4 for each participant group. The darker the shading, the greater the number of WoRs used. More than half of the BIN students routinely used two ways of reasoning. Almost half of the PIN students routinely used 3 WoRs (45 %), and almost all (92.5 %) used either 2, 3, or 4. More than half of the CT students (55 %) used 3 WoRs, and more than three fourths (85 %) used either 3 or 4. We found a statistical difference among the BIN, PIN, and CT groups, as determined by a one-way ANOVA ($F(2,116) = 30.9881, p < .0001$). A Tukey-Kramer post hoc test showed statistically significant differences between BIN and PIN students and BIN and CT students ($p < .01$), but not between PIN and CT students. Flexibility-across scores increased for comparisons of BIN to PIN and BIN to CT groups. One interpretation of these results is that school-based instruction or maturity or both contribute to flexibility development in the domain of integer arithmetic. Students in both the PIN and CT groups—7th graders and 11th graders who had already received instruction on computing with negative numbers—demonstrated greater flexibility-across problems than the BIN students, who had not yet experienced school-based instruction on additive integer arithmetic. That said, most BIN students, students who had yet to receive school-based instruction on additive integer arithmetic, demonstrated flexibility-across problems: Two-thirds used either two or three ways of reasoning.

Further, the percentage use of each WoR changed from the BIN to the PIN to the CT group. Fig. 2 depicts the overall percentage use of WoRs for students in each participant group. From BIN to PIN to CT, the use of Emergent (the WoR not included in the flexibility counts) decreased from 58 % to less than 1 %, whereas the use of Computational increased from 9 % to 75 %. Formal also increased across participant groups from 3 % to 24 %. This increase in Formal might be expected because students learn to generalize and leverage properties about the structure of our number system while they are expanding their numerical domains. The finding that participants continued to use a variety of WoRs across grade levels while use of particular WoRs ebbed and flowed is consistent with Siegler (1998, 2016) Overlapping Waves Metaphor “in which multiple ways of thinking coexist for prolonged periods, with development involving changes in their relative frequencies as well as introduction of new approaches” (p. 89, 1998).

6.2. Flexibility within

To illustrate students' flexibility-within problems, in Table 4, we share the percentages of participants who shared a secondary way of reasoning when prompted. Across these two open number sentences, about one fourth and two fifths, respectively, of the PIN and CT students were able to share a secondary WoR for both problems. However, more than one fourth of each of the PIN and CT groups were unable to share a secondary WoR for either problem. This latter finding is somewhat surprising given that, of these 13 students across the two groups, only 3 used just one WoR across the 25 problems. In other words, although most of these 13 students provided evidence that they knew and used multiple WoRs, none were able to bring forth an additional WoR for either problem when prompted.

6.3. Relationship Between Measures

6.3.1. Relationship between flexibility across and flexibility within

In Fig. 3, we share the average flexibility-within score in relation to the flexibility-across score. Two findings emerged. First, the students with the higher flexibility-across scores had, on average, higher flexibility-within scores (one-way ANOVA ($F(3,42) = 3.7644, p = .0176$)). Second, the average flexibility-within score among the 46 students was about half of the maximum score (1.07 on a 0–2 scale) whereas the average flexibility-across score was, on average, about three fourths of the maximum score (3.15 on a 0–4 scale). One possible interpretation for the relatively higher flexibility-across score is that flexibility across develops more readily than flexibility within. Another possible explanation is that the two problems for which a second WoR was requested do not lend themselves to multiple WoRs. However, $-2 - 7$ can be solved with an order-based approach or with a computational approach. Similarly, $-5 - -3$ can be readily solved with an analogy-based approach or with a computational approach. We think this latter hypothesis is an unlikely

Table 3
Measures.

Measure	Components Range of possible scores				
Flexibility across problems	Number of order-based 0–25	Number of analogy-based 0–25	Number of computational 0–25	Number of formal 0–25	WoRs used (3 + times) 0–4
Flexibility within problems	Ability to share a secondary WoR when prompted 0–2 –5 – – 3 no = 0, yes = 1				
Performance	Number (%) answered correctly 0–25 (0–100 %)				

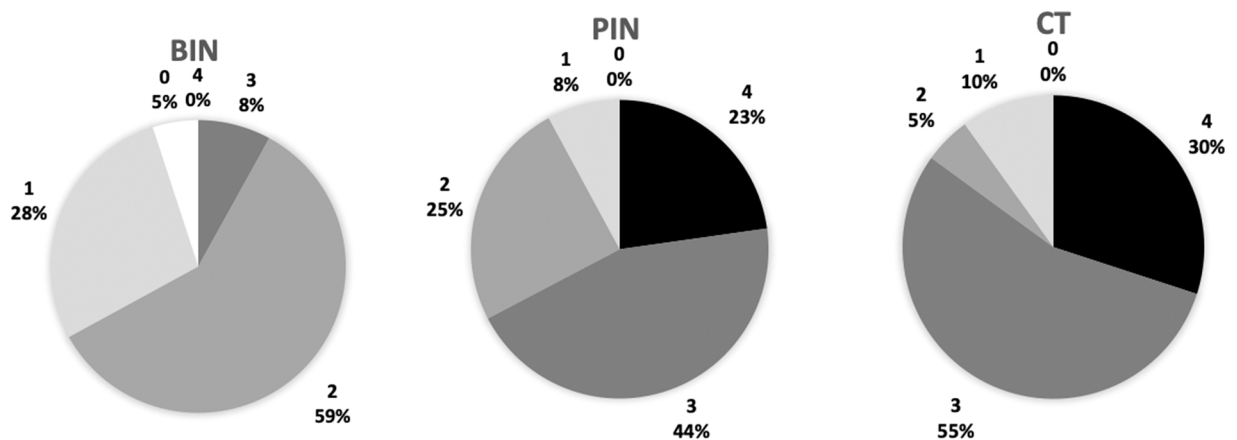
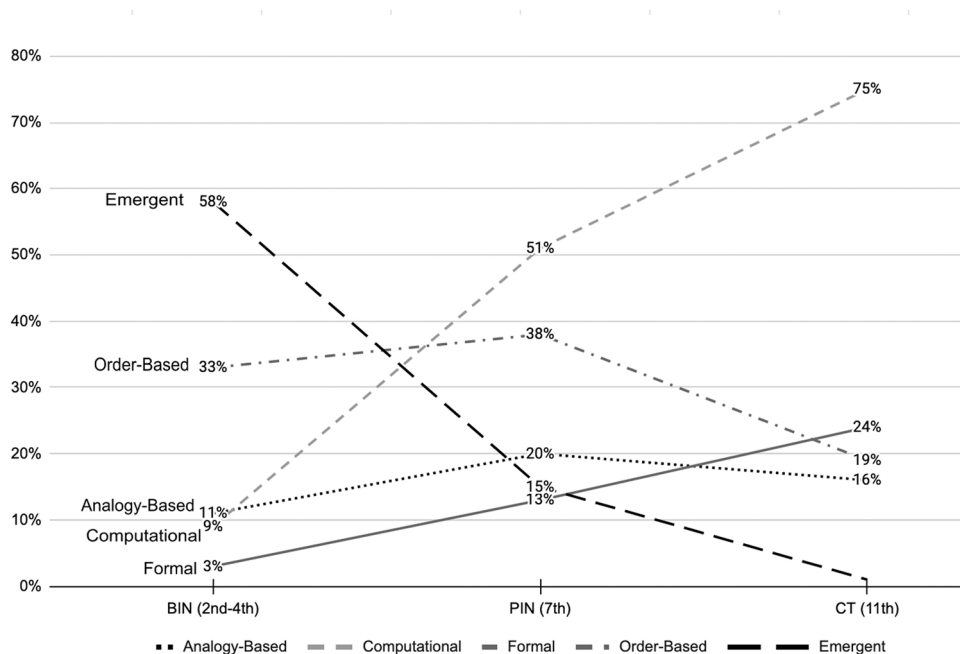
**Fig. 1.** Flexibility-across percentages by participant group. ¹Defined as the number of WoRs used 3 or more times (maximum score is 4).**Fig. 2.** Percentage use of ways of reasoning across groups.

Table 4
Percentage of Participants Who Shared a Secondary WoR When Prompted.

Percentage who shared a 2nd way on	PIN (<i>n</i> = 18)	CT (<i>n</i> = 28)
2 open number sentences	28 %	39 %
1 open number sentence	44 %	32 %
0 open number sentences	28 %	29 %

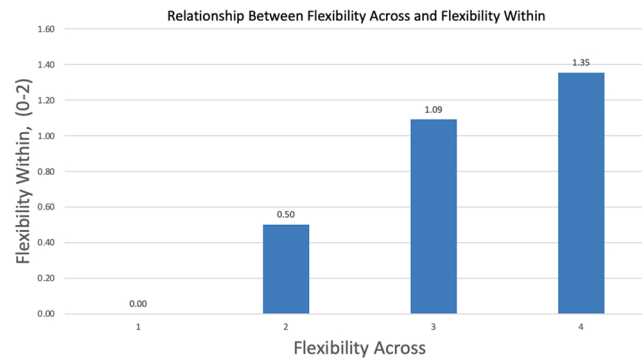


Fig. 3. Relationship between flexibility across and flexibility within.

Table 5
Correlations between flexibility and performance, one-tailed.

Measure	BIN	PIN	CT	Overall
Flexibility across	0.347* (<i>n</i> = 39)	0.523** (<i>n</i> = 40)	0.429** (<i>n</i> = 40)	0.662** (<i>n</i> = 119)
Flexibility within	—	0.237 (<i>n</i> = 18)	0.237 (<i>n</i> = 28)	0.173 (<i>n</i> = 46)

* $p < .05$.

** $p < .01$.

reason for these results. A third possible explanation is that we posed too few problems to too few students when examining flexibility within and so did not get the averages that we might have expected had we posed these prompts on more problems and to more students.

6.3.2. Relationship between flexibility and performance

Next, we investigated the degree to which flexibility and performance might be related, examining each of the subconstructs of flexibility across and flexibility within. The BIN students solved, on average, more than one third of the problems correctly (mean of 35.3 %), the PIN students solved about three fourths of the problems correctly (mean of 73.3 %), and CT students had a ceiling effect given that they solved virtually all problems correctly (mean of 98.2 %). We found that flexibility across was positively correlated with performance within participant groups and across groups (see Table 5).¹⁰ Students in the PIN group had the greatest correlation between flexibility across and performance. Additionally, the fact that flexibility across and performance were correlated regardless of whether participants have had school-based instruction is consistent with other research findings on flexibility. Finally, the correlations between flexibility within and performance were small and not statistically significant. There are several possible reasons for these results that we take up in the Limitations section.

6.4. Profiles

We now present profiles of one BIN, three PIN, and one CT student —Annie, Hannah, Sofia, Maria, and Carrie (pseudonyms), respectively—to exemplify differing ways that flexibility was exemplified across participant groups and to share a range in flexibility and performance.

¹⁰ As shared in 5.3.1.1.1, we previously reported correlations using a different measure of flexibility across (Lamb et al., 2018). In contrast to the previous operationalization of flexibility across, in the refined measure we share in this article, we consider students' preferences and competencies for using WoRs.

6.4.1. Profile of Annie

Annie, a 2nd grade BIN student, responded to 36 % of the problems correctly, and used two of the WoRs applied in our measure of flexibility. She used order-based reasoning on 32 % of the problems, analogy-based reasoning on 16 % of the problems, and emergent reasoning on 72 % of the problems; hence her flexibility across score was 2. Recall that the emergent category was not used in the flexibility calculation. Moreover, we note that responses could be assigned more than one WoR (see [Section 5.3](#)) so the percentage total of use of WoRs could be greater than 100 %, as in Annie's case. Additionally, Annie's performance varied depending on the problem type: she answered 83 % of the change-negative all-negative problems correctly but only 12.5 % of the change-negative not-all-negative problems correctly. She solved 43 % of the change-positive cross-zero problems correctly, but she answered every change-positive negative-side problem incorrectly.

For an example of an order-based approach that Annie used, she correctly answered "9" to the change-positive cross-zero problem, $3 - _ = -6$. Annie picked up a number line and placed her finger at 3 on the number line. Then, under her breath she counted while moving her finger one unit to the left at a time, "3, 2, 1, 0, -1, -2, -3, -4, -5, -6." She then placed one finger at -6 and then appeared to count silently the numbers between 3 and -6. She wrote "9" as her answer. She also used order-based reasoning for the problem $3 - 5 = _$ by counting back 5 units by ones on her fingers from 3 to get -2. Consistent with other BIN students, she used order-based reasoning on most change-positive problems, and was much more successful on change-positive cross-zero problems than on change-positive negative-side problems. For example, on the change-positive negative-side problem, $-8 - 3 = _$ she incorrectly answered, -5, using the order-based strategy of counting by ones, counting *up* from -8 to -5 (instead of counting *down* from -8 to -11).

Annie also used analogy-based reasoning. For example, for the change-negative all-negative problem $-5 + -1 = _$, Annie wrote -6 and answered, "negative six" as her answer, and as part of her explanation she shared, "It's just like the real numbers," comparing negative numbers to positive numbers (what Annie called "real numbers.") Similarly, for the change-negative all-negative problem, $-5 - -3 = _$, she put together 5 unifix cubes in a train. Then she removed 3 of the cubes from the train. She set down the 2 cubes remaining, wrote "-2," and shared aloud, "negative 2." She explained, "I took these (she picked up the 5 blocks). I knew it said three and I did that (removes three blocks from the five blocks). I knew it was two." The interviewer followed up, "You mentioned 5 and 3 and 2 but this is negative 5 and negative 3 and negative 2." Annie replied, "It was kind of the same but it wasn't. Like, it was kind of the same because they are the same numbers (points to the -5 and -3). It's just with the negatives in front of the 3 and the 5." Again, similar to other BIN students, she was successful on these types of problems, and answered 83 % of change-negative all-negative problems correctly.

When attempting to solve the change-negative not-all-negative problems $6 + _ = 4$ and $5 - _ = 8$, she shared that each answer was not possible, using emergent reasoning. For example, for $5 - _ = 8$, Annie shared "I don't think it's possible because this one [points to 5] is lower [than 8] so if you want to get to 8 you would have to put a plus. The interviewer asked, "So why doesn't minus [subtraction] work do you think?" Annie replied, "Because this one's [pointing to 8] a higher number and it wouldn't make sense because it's higher and not lower than five." Similar to other BIN students, Annie almost exclusively used emergent reasoning on the change-negative not-all-negative problems.

6.4.2. The profile of Hannah

Hannah, a PIN student, responded correctly to 32 % of the problems, and used two of the WoRs applied in our measure of flexibility. She used order-based reasoning on 84 % of the problems, computational reasoning on 12 % of the problems, and emergent reasoning on 64 % of the problems.

Although Hannah used order-based reasoning in conventional ways at times, more often than not her almost exclusive use of it (and perhaps the formulaic way in which she used this approach) hindered her success (cf. [Wessman-Enzinger, 2019a](#)). Hannah used order for change-positive problems by invoking the strategy *motion on a number line*. She correctly answered 71 % and 75 % of the change-positive cross-zero and change-positive negative side problems, respectively, and 0 % of the change-negative all-negative and change-negative not-all-negative problems. These former problems lend themselves to order-based reasoning. For example, for the problem $-3 + 6 = _$, Hannah placed her pen at -3 on the number line, moved to the right (because the problem was an addition problem) six numbers on the number line, counting by ones, and correctly landed on 3, which she shared as her answer. Similarly, for the problem $-8 - 3 = _$, a change-positive negative-side problem that only 60 % of the PIN participants answered correctly, Hannah placed her pen at -8, counted to the left three numbers on the number line, and correctly landed on -11. She shared as her answer "Negative 11." Hannah shared her reasoning, "Because I went to -8 [points to -8 on the number line] minus 3, so they took 3 more away, so negative 11." When prompted for how she knew which way to move on the number line, Hannah continued, "Because you were taking more away, so you go more negative."

However, Hannah also used the same order-based way of reasoning for almost every other problem and answered *every* one of these other problems incorrectly, predominantly because she continued to move right for addition and left for subtraction, irrespective of the sign of the change number (i.e., the second addend or the subtrahend). For example, for the change-negative all-negative problem $-5 + -1 = _$, Hannah placed her pen at -5 and (incorrectly) moved her pen to the right, wrote and stated as her answer "Negative 4," explaining her reasoning:

Because I went on my number line, and negative 5 plus (pointing to the addition sign in the problem), so [for] plus you have to go back up towards the positive numbers, and then just plus 1, and I got 4 (points to negative 4); I went to 4.

The interviewer followed up by pointing to the second addend of -1: "Do you think it mattered that it (this addend) was 1 or negative 1?" Hannah paused, looked down at her number line, and said, "No, I don't think it would [matter]." Additionally, for

subtraction problems, Hannah always moved left. For example, for the change-negative not-all-negative problem $6 - - 2 = _$, Hannah placed her pen at 6 on the number line and paused. She then moved her pen to the left two units and answered 4, explaining, "You go to positive 6 minus negative 2 equals 4." The interviewer asked for clarification, and Hannah added, "I am just a little bit confused by this (points to the open number sentence), so I just looked at this (points to the number line). So, then I just minused 2, and I got 4." The discussion continued.

Interviewer: Which part is a little confusing?

Hannah: Because it is a positive 6 minus a negative 2 (she points to the number line). I don't know if you go down here (she moves her pen to the negative side of the number line), or, I don't know, but I think it's 4."

Interviewer: Was there something else (another answer) you were considering?

Hannah: Negative 4, but then I was like, "No. Because it $[- 4]$ is all the way down here (points to the negative side of the number line)."

Hannah appeared to implicitly recognize that the problem had a structure different from the other problems she had successfully solved, and she shared her confusion about whether or how to adapt her use of motion on the number line when subtracting a negative number.

Hannah's nearly exclusive use of order-based reasoning seems to have limited her options for solving problems, especially change-negative problems. She was often unsuccessful, as were most others who predominantly relied on one way of reasoning. In the next profile, we describe Sofia. Like Hannah, Sofia tended to rely primarily on one way of reasoning and also had limited success. Unlike Hannah, however, Sofia primarily used computational reasoning.

6.4.3. The profile of Sofia

Sofia was more successful than Hannah, inasmuch as she answered about two thirds (64 %) of the problems correctly. She used two ways of reasoning: a computational way of reasoning on about two thirds (68 %) of all open number sentences and order-based reasoning on about one third (32 %). Sofia's success was also related to problem type but the problem types she successfully navigated were different from those that Hannah correctly solved. Sofia correctly answered 88 % of the change-negative not-all-negative problems, problems that many PIN students struggled to answer. However, she correctly answered only 50 % of the change-negative all-negative problems, 25 % of the change-positive negative-side problems, and 71 % of the change-positive cross-zero problems.

For an example of the computational WoR she used on a change-negative not-all-negative problem, $5 - _ = 8$, Sofia correctly answered $- 3$, saying, "If it's negative and negative (pointing to the subtraction sign and the negative sign on $- 3$), you change the sign, ... both signs, and it will become a positive, and it will equal 8." For this problem, Sofia invoked the *Keep-Change-Change* (KCC) rule described in Section 2.1. Similarly, for the problem $6 + _ = 4$, Sofia correctly answered $- 2$, using a different rule. In particular, Sofia used what is often referred to as the *different-signs rule*, such that when adding a negative number and a positive number, one finds the difference of the absolute values of the numbers and appends the sign of the number that has the larger absolute value.

Sofia's focus on primarily computational approaches likely explains why she also applied the different-signs rule on problems for which the rule was not applicable. For example, for the problem $- 8 - 3 = _$, Sofia incorrectly answered $- 5$, explaining that she subtracted 3 from 8 and, because the 8 was negative, her answer should also be negative. "When [the absolute value of] a negative number is bigger than a 3, like, a positive number (points to 3), if this one (points again to 3) is lower [than the absolute value of $- 8$], then the answer becomes a negative sign." Here, Sofia invoked the different-signs rule on a subtraction problem, a problem for which the rule is invalid. Similarly, Sofia inappropriately invoked a rule for multiplying negative numbers to complete $- 5 + - 1 = _$. Sofia explained, "I think it is 6 because if you add both negatives, then it becomes a (negative) 6, but if you see the signs negative and a negative, it becomes a positive number, so I say it's a 6." When asked why two negatives become a positive, Sofia responded, "Because when the signs are the same, it becomes a positive. Well, my teacher said that if. it was two negatives, then it [the result] would become a positive." In this example, Sofia used another rule that she knew—that two negatives, when multiplied, result in a positive product—and misapplied it on an addition problem with two negative numbers.

Sofia's computational way of reasoning appeared to override attempts to look for alternative approaches, even when she was prompted, as the following example shows. Sofia successfully used the Keep-Change-Change rule for the all-negative problem $- 5 - - 3 = _$, correctly answering $- 2$. Wondering whether Sofia could solve the problem as it was originally written, the interviewer asked if she could solve $- 5 - - 3 = _$. Sofia shared, "I think it would be (pauses). Well I don't know because if you don't change it (the problem to $- 5 + + 3$), then it would become like, (whispers) negative. (She pauses again and looks at the interviewer.) I don't know." Although Sofia was able to correctly solve the problem, she was unable to share another way of reasoning, even when prompted. Additionally, Sofia's orientation toward using rules appeared to hinder an orientation toward trying to make sense of the problem in another way.

Like Hannah's predominant use of order-based reasoning, Sofia's almost exclusive use of computational reasoning appeared to hamper her success. Additionally, when prompted, she was unable to generate a secondary WoR different from computational. Looking across both Hannah's and Sofia's profiles, we see that the particular way of reasoning they used was less important than the fact that each student appeared to have primarily *one* way of reasoning on which she relied. And being limited to a single way of reasoning appeared to negatively influence their success.

6.4.4. The profile of Maria

In contrast to Hannah and Sofia, Maria responded correctly to every problem and she used all four ways of reasoning. Maria invoked every one of the four ways of reasoning, using computational reasoning on 48 % of the problems, order-based reasoning on 44 % of the problems, analogy-based reasoning on 28 % of the problems, and formal reasoning on 24 % of the problems. In examples of

the range in her use of WoRs, Maria used analogy-based reasoning on the problem $-5 + -1 = _$ by comparing negative numbers to “bad guys.” For this problem she correctly answered -6 , saying, “Since negative numbers are like bad guys, 5 bad guys met up with 1 more bad guy, so there were 6 bad guys total.” She used order-based reasoning for $-3 + 6 = _$ by decomposing 6 into 3 and 3 so she could “jump to 0” using a partial sum. Maria explained her answer of 3 saying, “Half of 6 is 3, so then that would bring it to the 0. And 3 more would bring it to the 3. And that would equal 6.” Her strategy of using a decade number of 0 supported her to count more efficiently than counting by 1 s. In contrast, she correctly completed the number sentence $6 - 2 = _$ using computational reasoning, explaining, “I changed the signs so it was plus-plus, and $6 + 2$ is 8.” Finally, Maria used formal reasoning for $6 + _ = 4$. She answered, “Negative 2. It’s [the unknown is] not going to be a positive because the sum is less than 6 and it [the operation] is addition. So, it [the missing addend] has to be a negative number. So, -2 .” Maria used deductive reasoning to determine the sign of the unknown by noticing that if she added two addends and the sum was less than one of the addends, then the other addend must be a negative number. Her response is formal reasoning because she uses the indicated operation and the relative sizes of the numbers to deduce the unknown sign on the basis of generalizations she has made about how addition and subtraction work when number systems are extended (from whole numbers to integers in this case). Further, for the problems $-5 - -3$ and $-2 - 7$ Maria shared secondary WoRs when asked for another strategy. In both cases, she initially used a combination of computational and analogy-based reasoning, and her secondary WoR was order-based. Her robust flexibility across problems was a hallmark of her approach to solving these problems.

Similar to other students who used all four ways of reasoning, Maria tended to invoke those ways of reasoning that corresponded to the underlying structure of the problem types. For example, she used order-based reasoning on 71 % of the change-positive cross-zero problems, and either computational, formal, or both kinds of reasoning on 88 % of the change-negative not-all-negative problems, indicating that her (implicit) attention to the underlying structure evoked particular ways of reasoning. This attention appeared to be influential in the way of reasoning she used and, presumably, in her success.

6.4.5. The profile of Carrie

CT-student Carrie responded correctly to every problem and she used all four ways of reasoning. Similar to other CT students, she confidently and quickly responded to problems. She used computational or formal reasoning or both on every change-negative not-all-negative problems, order-based reasoning on 75 % of change-positive negative-side problems but on only 14 % of change-positive cross-zero problems, and used analogy-based reasoning on 50 % of the change-negative all-negative problems.

For an example of order-based reasoning, to solve the change-positive negative-side problem $-8 - 3 = _$, Carrie gave the correct answer of -11 , and shared her reasoning, “Negative eight and then if you go three more to the left, then it’s negative 11.” The interviewer followed up, “Do you mean like on the number line when you say [to the left]?” Carrie nodded her head affirmatively. Although Carrie did not physically use a number line to get her answer, she indicated to the interviewer that she was referring to a number line (possibly imagining one in her mind) when sharing that “you go three more to the left.” For the change-negative all-negative problem, $-5 + -1 = _$, Carrie answered -6 and used analogy-based reasoning when she shared, “because five plus one is six and they’re both negative. So your answer is just gonna be the same thing (as with positive numbers) but negative.” Similar to Maria, Carrie used computational reasoning (KCC) to solve $6 - -2 = _$ and formal reasoning to solve $6 + _ = 4$.

6.4.6. Looking across profiles

The five students exhibited a range in flexibility and approaches to solving problems. For example, whereas BIN-student Annie never transformed a problem into a new problem by using a computational strategy, PIN-students Sofia and Maria and CT-student Carrie transformed many problems, often invoking KCC. Further, Annie successfully used analogy-based reasoning on change-negative all-negative problems, whether they were addition or subtraction problems, but Hannah and Sofia never flexibly used analogy-based reasoning, and Maria and Carrie used it only on change-negative all-negative problems involving addition, such as $-5 + -1$. Moreover, Annie and the BIN students were more likely to use concrete representations (cubes, the physical number line, or fingers) to aid their reasoning, using concrete representations an average of 5.7 times per student. In contrast, the CT students such as Carrie were less likely to use any concrete representations, instead making references to visualizing a number line in their heads as they were solving: the CT students used concrete representations an average of .7 times per student. Finally, Annie, like others in the BIN group, used emergent reasoning on many problems, whereas the use of emergent reasoning was extremely rare with the CT students; Carrie never used it. We further discuss the profiles of Hannah and Sofia in the Discussion.

7. Discussion

In this article, we investigated flexibility in students’ Ways of Reasoning about additive integer arithmetic. We documented statistically significant differences across the BIN, PIN, and CT groups in relation to flexibility across by documenting the average number of Ways of Reasoning used, showing a statistically significant relationship between flexibility-within scores and flexibility-across scores, and identifying moderate-to-strong correlations between flexibility across and performance on integers open number sentences, within and across participant groups. Additionally, the profiles of 2nd grader Annie, 7th graders Hannah, Sofia, and Maria, and 11th grader Carrie provided insight into how flexibility may influence WoR choice and, ultimately, performance.

We extended the research on flexibility by describing and measuring flexibility in a content domain that has received less attention, replicating findings of the positive correlation between flexibility across and performance that has been shown in many other domains, and categorizing ways that researchers distinguish flexibility by introducing the terms *flexibility within* and *flexibility across*. Additionally, consistent with others (e.g., Lemaire & Siegler, 1995; Siegler, 1987), we documented that students, prior to school-based instruction, demonstrate flexibility across problems. In the next sections, we discuss (a) the flexibility of students prior to

school-based instruction, (b) the profiles of the PIN students with low flexibility, (c) the benefits of each component of flexibility, and (d) the cross-sectional nature of our findings.

7.1. Assessing flexibility prior to school-based instruction

Siegler (1987) and Lemaire and Siegler (1995) assessed students' strategy choices for subtraction (Grades K-1) and multiplication (Grade 2). By including the BIN participant group in our study, we contributed to the flexibility literature by showing that students in grades 2 and 4, demonstrate flexibility across problems in additive integer reasoning. Specifically, two thirds of those 2nd and 4th graders who had knowledge of negative integers used at least two WoRs across the 25 open number sentences. We view this finding as an indication that young learners possess an implicit desire to make sense that manifests in their implicitly choosing ways of reasoning depending on the numbers in the problem, the perceived difficulty of the problem, and the children's competence with each WoR. This finding may help us, as a field, understand how to better foster and develop flexibility by building on students' intuitive uses of it. Additionally, we shared the profile of 2nd grader Annie to reveal details about the ways in which many young students solved particular problem types prior to instruction. For example, on change-positive problem types, Annie tended to use order-based reasoning by counting up or back by ones on her fingers or using the number line. She tended to use analogy-based reasoning with cubes and comparisons to "real" (whole) numbers to solve change-negative-all-negative problems, and she primarily used emergent reasoning to solve change-negative-not-all-negative problems such as $6 + _ = 4$, declaring most of these problems "not possible" to solve. Documenting these approaches provides researchers and teachers starting points for connecting with young students' ideas when solving different problem types.

7.2. Low flexibility across scores: the profiles of Hannah and Sofia

Returning to the profiles of Hannah and Sofia, we wondered whether Hannah's and Sofia's struggles (and those of others who fit the profile of having low flexibility and limited performance) were due to their primarily having one WoR they used or to their seeming use of that one WoR in a procedural way, without necessarily considering whether their approach made sense for a given problem. We believe that both conjectures are likely true. Theoretically, if a student could use one WoR sensibly, then the student might be able to solve many problems correctly *and* also be able to realize when the chosen WoR is less useful than another for some problems. However, we suggest that a rigid devotion to a particular WoR is likely to inhibit sensemaking, as evidenced in the cases of Hannah and Sofia and as seen in others' research (Newton et al., 2010; Rittle-Johnson et al., 2012).

Further, the profiles of Hannah and Sofia—students who primarily used one WoR—also offer starting points for helping students with similar profiles to increase their flexibility and understanding in this domain. By looking beyond the overall percentage correct, we can begin to imagine ways to provide support. For example, by attending to which problem types students more and less successfully solved along with their predominant WoRs, we can begin to offer targeted interventions that might have a greater chance of success than a one-size-fits-all approach. For example, knowing that Hannah could successfully solve most change-positive problems using order-based reasoning, one next step could be for her to compare the similarities and differences between change-positive and change-negative problems, such as comparing $-2 + 6 = _$, and $-2 + -6 = _$. She may experience some cognitive disequilibrium that allows her to consider other approaches, or to consider how she could adapt her order-based approach along with formal reasoning to solve the change-negative problem. Or, knowing that Sofia successfully used two different computational WoRs, KCC and the different-signs rule, for the change-negative not-all-negative problems, she could be asked to share her approaches with the class. Then, another student could share an order-based WoR for a change-positive cross-zero problem and they could discuss the features of the problems that lend one approach to making one type of problem easier to solve than another. In this way, teachers can tailor their instruction by building on what students already know and supporting students' growth from there.

7.3. Two related components of flexibility

We contributed to the flexibility literature by measuring both components of flexibility in the same study, similar to Blöte et al. (2001) and Berk et al. (2009). We further contributed to the conceptualization of flexibility by naming two related, but distinct, constructs and including them under the umbrella of flexibility—flexibility-within problems and flexibility-across problems. We documented their relationship to each other within the domain of additive integer arithmetic and showed a change in flexibility over time through a cross-sectional analysis. We believe both types of flexibility are valuable for students. Flexibility across benefits students because differences in problem characteristics make problems easier to solve with certain WoRs. Flexibility within benefits students because having more than one way to solve a problem provides students with an alternative approach if an initial WoR is difficult to apply for a given problem. Additionally, flexibility within provides students with a different kind of image about the nature of mathematics. By recognizing that the same problem can be solved in many ways, students experience the creative and innovative aspects of mathematics, aspects that may be overlooked in school mathematics.

7.4. Assessing flexibility at three time points

Consistent with Siegler's Overlapping Waves Metaphor (1998, 2016), our findings show that WoRs are not phased out while students mature. That is, students across grade levels continued to use multiple Ways of Reasoning to solve additive integer problems; the use of some strategies decreased, and use of others increased (see Fig. 2). The results may indicate that developing flexibility with

additive integer arithmetic may initially involve developing and using multiple WoRs and, later, refining the ability to use WoRs that are better suited for some problem types than others. The finding that college-track 11th graders were both the most flexible in their ways of reasoning and the most accurate shows that particular ways of reasoning are not replaced by other, more sophisticated, ways of reasoning but, rather, that students who can fluently solve these number sentences appear to become more sophisticated in (implicitly) selecting which WoRs they use, a finding consistent with [Siegler \(1998\)](#). In other words, flexibility is key. We believe that the growth in flexibility can support students to develop deeper conceptions of integers, to choose WoRs that are better targeted to problem types, and to improve their accuracy.

8. Implications

These findings, taken alongside research about the benefits of intentionally supporting the development of flexibility in other content domains, provide implications for practice. We suggest that the development of flexibility in students' reasoning about additive integer arithmetic should be an explicit curricular and instructional goal. Further, we conjecture that flexibility across develops before flexibility within, and so we suggest that teachers support students by asking them to compare strategies across *different* (in addition to the same) additive integer problems. In [Sections 8.1](#) and [8.2](#), we provide a rationale for each implication.

8.1. Developing flexibility about additive integer arithmetic as an explicit instructional goal

One goal of integer instruction should be for students to develop multiple WoRs and the flexibility to use them when solving various types of additive integer problems. We suggest that curricula and pedagogy should be designed to leverage students' implicit desires for flexibility. However, we know that some teachers discourage flexibility in favor of more general, one-approach-fits-all methods for completing additive integer arithmetic tasks ([Hawthorne et al., 2022](#)).

We found that successful students tended not to use a single way of reasoning on every (or almost every) problem. Thus, teaching students one all-encompassing way may have the unintended consequence of impeding students' success by limiting their flexibility (cf. [Stephan & Akyuz, 2018](#)). Further, if teachers attempt to teach a single way of reasoning and that way of reasoning does not make sense to the child, the repercussions may extend beyond performance with additive integer arithmetic. That approach may have the unintended consequence of negatively influencing beliefs about mathematics and the nature of mathematics, students' mathematical identities, and so on. Speaking to a similar conclusion more than 25 years ago and still relevant today, [Smith \(1995\)](#) wrote,

A serious danger of [following a traditional textbook-based] approach is its failure to recognize explicitly students' specific and constructed strategies as legitimate mathematical knowledge. Insightful student strategies are often criticized and rejected only because they fail to match the textbook methods. Over time, students learn that mathematics is about remembering general numerical procedures and has little to do with their own creative thinking and understanding. (p. 41)

Evidence from research, ours and others' ([Bofferding, 2014](#); [Wessman-Enzinger, 2019b](#)), indicates that young students, prior to instruction, demonstrate rich ways of reasoning from which teachers could build. Many researchers who have attempted to increase students' flexibility have typically done so *after* students have already learned standard approaches (see [Blöte et al., 2001](#) for an investigation designed to develop flexibility before and after instruction). If teachers can nurture and support students to develop flexibility *before* or at the same time that standard approaches are learned, we conjecture that the students' development of flexibility may be less daunting. An instructional focus on flexibility has the potential to provide students with greater access to mathematics, develop their beliefs about mathematics as being a creative endeavor, and improve performance.

8.2. Prompting students to compare strategies for solving different (in addition to the same) problems

On the bases of our analyses within integer arithmetic, we conjecture that students may develop flexibility-across problems before they develop flexibility-within problems. That is, asking students for a second way to solve a single open number sentence may be a useful prompt, and comparing ways that students have used to solve different problems and discussing why one way is more or less useful for a particular problem may be a better starting point. If we expect students to use a secondary WoR within a single problem as a way to develop or promote flexibility, we may be missing a critical component of the developmental progression. Students may *first* develop flexibility-across problems by implicitly recognizing ways of reasoning that align with problem features, and students later develop flexibility-within problems by analyzing the same problem from multiple perspectives and then considering how to use a given WoR for a given problem. Thus, expecting students to have multiple WoRs to solve a single problem before they have had opportunities to develop WoRs across problems may be bypassing opportunities for students to develop flexibility, and, by extension, to deepen their understanding of the mathematics.

Taking into account successful instructional interventions for supporting flexibility in other content domains, (e.g., [Berk et al., 2009](#); [Rittle-Johnson & Star, 2007](#)), we suggest that teachers have students compare and contrast ways of reasoning used on problems of different types because specific problem types are more likely to evoke one way of reasoning than another. For example, teachers might have students compare ways of reasoning for a change-negative all-negative problem to ways of reasoning used for a change-positive cross-zero problem to allow students to identify features of each problem that made one way of reasoning easier for solving one problem type than another (cf. [Bofferding & Aqazade, 2018](#)). As identified in previous studies of flexibility, focusing on a single WoR rather than comparing different ways can inhibit the development of flexibility, and so comparing WoRs in the domain of additive integer arithmetic seems promising. Supporting teachers to foster their students' development of flexibility in this domain is

crucial. Finally, we believe that flexibility, in general, is an important component of mathematical proficiency that is often overlooked. We suggest that an emphasis on flexibility in integer arithmetic is but one part of a broader instructional approach in which development of flexibility is a goal across the mathematics landscape.

9. Limitations and next steps

Although we learned much about students' flexibility in the domain of additive integer reasoning, we have more to do. For example, we conducted a secondary analysis related to flexibility as part of a large research study. By only looking at two tasks, we may not have had the sensitivity to get an accurate measure of our participants' flexibility-within. Further, smaller samples tend to yield noisier correlations, and it could be the case that our correlations were lower than the true correlation (Goodwin & Leech, 2006). In a subsequent study, we would explicitly pose flexibility-within problems to every participant. That is, we would ask every child to solve multiple problems in more than one way, rather than asking for a second way of solving only for a subset of our participants.

Further, as a field we need to investigate the extent to which providing students opportunities to develop and reflect on many ways of reasoning actually improves flexibility and performance in the realm of additive integer arithmetic. Can we reasonably expect that improved flexibility will be associated with improved performance? We believe so. Studies focused on supporting flexibility in other content areas have routinely shown that an instructional focus on flexibility improves both flexibility and performance (e.g., Berk et al., 2009; Blöte et al., 2001; Rittle-Johnson & Star, 2007). We suggest that one next step would be to investigate instructional approaches that support the development of students' flexibility in the context of additive integer arithmetic. We are optimistic that such a focus on flexibility could not only provide students with greater access to the mathematics and improved performance but could also improve their mathematical identities and support the notion of mathematics as a creative and worthwhile endeavor with human inventiveness at its core.

Funding

This material is based on work supported by the National Science Foundation (NSF) under grant number 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.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Thank you to Bonnie Schappelle who edited several drafts of this manuscript.

References

- Aqazade, M., & Bofferding, L. (2021). Second and fifth graders' use of knowledge-pieces and knowledge-structures when solving integer addition problems. *Journal of Numerical Cognition*, 7(2), 82–103.
- 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. <https://doi.org/10.1080/10986060903022714>
- Bishop, J. P., Lamb, L. L., Philipp, R. A., Whitacre, I., & Schappelle, B. (2022). Beyond the sign rules: Problem types and ways of reasoning for addition and subtraction of integers. *Mathematics Teacher: Learning and Teaching Pre-K–12*, 115(3), 202–210.
- Bishop, J. P., Lamb, L. L., Philipp, R. A., Whitacre, I., & Schappelle, B. P. (2014b). Using order to reason about negative numbers: The case of violet. *Educational Studies in Mathematics*, 86(1), 39–59.
- Bishop, J. P., Lamb, L. L., Philipp, R. A., Whitacre, I., & Schappelle, B. P. (2016). Leveraging structure: Logical necessity in the context of integer arithmetic. *Mathematical Thinking and Learning*, 18(3), 209–232.
- Bishop, J. P., Lamb, L. L., Philipp, R. A., Whitacre, I., & Schappelle, B. P. (2018). 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* (pp. 47–71). Switzerland: Springer.
- Bishop, J. P., Lamb, L. L., Philipp, R. A., Whitacre, I., Schappelle, B. P., & Lewis, M. (2014a). 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.
- 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.
- Bofferding, L. (2014). Negative integer understanding: Characterizing first graders' mental models. *Journal for Research in Mathematics Education*, 45(2), 194–245.
- Bofferding, L. (2019). Understanding negative numbers. In *Constructing number* (pp. 251–277). Cham, Switzerland: Springer, Cham.
- Bofferding, L., Aqazade, M., & Farmer, S. (2018). Playing with integer concepts: A quest for structure. In *Exploring the integer addition and subtraction landscape* (pp. 3–23). Cham, Switzerland: Springer, Cham.
- 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: University of Illinois at Chicago.
- Bofferding, L., & Wessman-Enzinger, N. (2017). Subtraction involving negative numbers: Connecting to whole number reasoning. *The Mathematics Enthusiast*, 14(1), 241–262. <https://doi.org/10.54870/1551-3440.1396>
- Bofferding, L., & Aqazade, M. (2018). Second and fifth graders' integer subtraction performance: Learning from contrasting worked example. In T. E. Hodges, G. J. Roy, & A. M. Tyminski (Eds.), (2018). *Proceedings of the 40th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 183–186). Greenville, SC: University of South Carolina & Clemson University.

- Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (2014). *Children's mathematics: Cognitively Guided Instruction*. Portsmouth, NH: Heinemann.
- Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loei, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, 26(4), 499–531.
- Carpenter, T. P., Levi, L., Berman, P. W., & Pligge, M. (2005). Developing algebraic reasoning in the elementary school. In *Understanding mathematics and science matters* (pp. 81–98). Routledge.
- Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, 15(3), 179–202.
- Carpenter, T. P., Moser, J. M., & Hiebert, J. (1981). Problem structure and first-grade children's initial solution processes for simple addition and subtraction problems. *Journal for Research in Mathematics Education*, 12(1), 27–39.
- Clements, D. H., Swaminathan, S., Hannibal, M. A. Z., & Sarama, J. (1999). Young children's concepts of shape. *Journal for Research in Mathematics Education*, 192–212.
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research*. Sage Publications.
- Dowker, A. D., Flood, A., Griffiths, H., Harriss, L., & Hook, L. (1996). Estimation strategies of four groups. *Mathematical Cognition*, 2(2), 113–135. <https://doi.org/10.1080/135467996387499>
- 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.
- Franke, M. L., Webb, N. M., Chan, A. G., Ing, M., Freund, D., & Battey, D. (2009). Teacher questioning to elicit students' mathematical thinking in elementary school classrooms. *Journal of Teacher Education*, 60(4), 380–392.
- Gallardo, A. (2002). The extension of the natural-number domain to the integers in the transition from arithmetic to algebra. *Educational Studies in Mathematics*, 49(2), 171–192. <https://doi.org/10.1023/A:1016210906658>
- Ginsburg, H. P. (1997). *Entering the child's mind: The clinical interview in psychological research and practice*. New York: Cambridge University Press.
- Glancy, A. W., & Pettis, C. (2017). Analysis of the relative difficulty of different integer problem types. In E. Galindo & J. Newton (Eds.), *Proceedings of the 39th annual meeting of the North American Chapter of the international group for the psychology of mathematics education* (pp. 275–278). Indianapolis, IN: Hoosier Association of Mathematics Teacher Educators.
- Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r . *Journal of Experimental Education*, 74(3), 249–266.
- Hästö, P., Palkki, R., Tuomela, D., & Star, J. R. (2019). Relationship between mathematical flexibility and success in national examinations. *European Journal of Science and Mathematics Education*, 7(1), 1–13.
- Hawthorne, C., Philipp, R. A., Lamb, L. L., Bishop, J. P., Whitacre, I., & Schappelle, B. P. (2022). Reconceptualizing a mathematical domain on the basis of student reasoning: Considering teachers' perspectives about integers. *The Journal of Mathematical Behavior*, 65, Article 100931. <https://doi.org/10.1016/j.jmathb.2021.100931>
- Hefendehl-Hebeker, L. (1991). Negative numbers: Obstacles in their evolution from intuitive to intellectual constructs. *For the Learning of Mathematics*, 11(1), 26–32.
- Heirdsfield, A. M., & Cooper, T. J. (2002). Flexibility and inflexibility in accurate mental addition and subtraction: Two case studies. *Journal of Mathematical Behavior*, 21(1), 57–74.
- Herbst, P., Chazan, D., Crespo, S., Matthews, P. G., & Lichtenstein, E. K. (2022). How manuscripts can contribute to research on mathematics education: An expansive look at basic research in our field. *Journal for Research in Mathematics Education*, 53(1), 2–9.
- 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.
- Lamb, L. L., Bishop, J. P., Philipp, R. A., Whitacre, I., & Schappelle, B. P. (2018). A cross-sectional investigation of students' reasoning about integer addition and subtraction: Ways of reasoning, problem types, and flexibility. *Journal for Research in Mathematics Education*, 49(5), 575–613.
- Le Fevre, 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. <https://doi.org/10.1037/0096-3445.125.3.284>
- Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, 124(1), 83–97.
- 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. <https://doi.org/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.
- Newton, K. J., Star, J. R., & Lynch, K. (2010). Understanding the development of flexibility in struggling algebra students. *Mathematical Thinking and Learning*, 12(4), 282–305. <https://doi.org/10.1080/10986065.2010.482150>
- Nistal, A. A., Van Dooren, W., & Verschaffel, L. (2012). What counts as a flexible representational choice? An evaluation of students' representational choices to solve linear function problems. *Instructional Science*, 40(6), 999–1019.
- Nistal, A. A., Van Dooren, W., & Verschaffel, L. (2014). Improving students' representational flexibility in linear-function problems: An intervention. *Educational Psychology*, 34(6), 763–786. <https://doi.org/10.1080/01443410.2013.785064>
- Numberger-Haag, J. (2018). Take it away or walk the other way? Finding positive solutions for integer subtraction. In *Exploring the integer addition and subtraction landscape* (pp. 109–141). Cham, Switzerland: Springer, Cham.
- 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.
- Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, 99(3), 561.
- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2012). Developing procedural flexibility: Are novices prepared to learn from comparing procedures? *British Journal of Educational Psychology*, 82(3), 436–455.
- Shaw, S. T., Pogossian, A. A., & Ramirez, G. (2020). The mathematical flexibility of college students: The role of cognitive and affective factors. *British Journal of Educational Psychology*, 90, 981.
- Siegler, R. S. (1987). Strategy choices in subtraction. In J. Sloboda, & D. Rogers (Eds.), *Cognitive process in mathematics* (pp. 81–106). New York: Oxford University Press.
- Siegler, R. S. (1998). Strategic developing: Trudging up the staircase of swimming with the tide? In *Emerging minds: The process of change in children's thinking*, 84–102. Oxford University Press, Incorporated. ProQuest Ebook Central. (<http://ebookcentral.proquest.com/lib/csla/detail.action?docID=271379>).
- Siegler, R. S. (2016). Continuity and change in the field of cognitive development and in the perspectives of one cognitive developmentalist. *Child Development Perspectives*, 10(2), 128–133.
- Siegler, R. S., Adolph, K. E., & Lemaire, P. (1996). Strategy choices across the life span. In *Implicit memory and metacognition* (pp. 79–121). Mahwah, NJ: Erlbaum.
- Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, 126(1), 71–92.
- Smith, J. P., III (1995). Competent reasoning with rational numbers. *Cognition and Instruction*, 13(1), 3–50. https://doi.org/10.1207/s1532690xcil301_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.
- Star, J. R., & Newton, K. J. (2009). The nature and development of experts' strategy flexibility for solving equations. *Zentralblatt für Didaktik der Mathematik*, 41(5), 557–567.
- Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, 18(6), 565–579. <https://doi.org/10.1016/j.learninstruc.2007.09.018>

- Star, J. R., Tuomela, D., Prieto, N. J., Hästö, P., Palkki, R., Abánades, M.Á., & Liu, R. D. (2022). Exploring students' procedural flexibility in three countries. *International Journal of STEM Education*, 9(1), 1–18.
- Stephan, M., & Akyuz, D. (2012). A proposed instructional theory for integer addition and subtraction. *Journal for Research in Mathematics Education*, 43(4), 428–464. <https://doi.org/10.5951/jresmetheduc.43.4.0428>
- Stephan, M., & Akyuz, D. (2018). Teaching integers to students with disabilities: Three case studies. In *Exploring the integer addition and subtraction landscape* (pp. 75–108). Cham: Switzerland: Springer, Cham.
- 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. <https://doi.org/10.1007/s10649-006-9077-6>
- Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, 24(3), 335–359.
- 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. <https://doi.org/10.1023/A:1020229023965s>
- Wessman-Enzinger, N. M. (2018). Integer play and playing with integers. In *Exploring the integer addition and subtraction landscape* (pp. 25–46). Cham, Switzerland: Springer, Cham.
- Wessman-Enzinger, N. M. (2019aaa). Grade 5 children's drawings for integer addition and subtraction open number sentences. *The Journal of Mathematical Behavior*, 53, 105–128.
- Wessman-Enzinger, N. M. (2019bbb). Integers as directed quantities. In A. Norton, & M. W. Alibali (Eds.), *Constructing number: Merging perspectives from psychology and mathematics education* (pp. 279–305). Cham, Switzerland: Springer, Cham.
- Wessman-Enzinger, N. M., & Tobias, J. M. (2022). The dimensions of prospective elementary and middle school teachers' problem posing for integer addition and subtraction. *Journal of Mathematics Teacher Education*, 25(1), 1–33. <https://doi.org/10.1007/s10857-020-09477-x>
- Wessman-Enzinger, N.M. (2015). Alice's drawings for integer addition and subtraction open number sentences. In T.G. Bartell, K.N. Bieda, R.T. Putnam, K. Bradfield, & H. Dominguez (Eds.), *Proceedings of the 37th annual meeting of the North American chapter of the international group for the psychology of mathematics education* (241–244). East Lansing, MI: Michigan State University.
- Xu, L., Liu, R. D., Star, J. R., Wang, J., Liu, Y., & Zhen, R. (2017). Measures of potential flexibility and practical flexibility in equation solving. *Frontiers in Psychology*, 8 (1368), 1–13.