Journal of Mathematical Behavior 45 (2017) 47–62

Contents lists available at ScienceDirect

The Journal of Mathematical Behavior

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

Integer comparisons across the grades: Students' justifications and ways of reasoning☆

Ian Whitacre[a,*], Beti Azuz[b], Lisa L.C. Lamb[b], Jessica Pierson Bishop[c], Bonnie P. Schappelle[b], Randolph A. Philipp[b]

[a] Florida State University, United States
[b] San Diego State University, United States
[c] Texas State University, United States

ARTICLE INFO

Article history:
Received 20 March 2016
Received in revised form
15 November 2016
Accepted 20 November 2016
Available online 11 December 2016

Keywords:
Integers
Negative numbers
Children's mathematical thinking
Order
Magnitude

ABSTRACT

This study is an investigation of students' reasoning about integer comparisons—a topic that is often counterintuitive for students because negative numbers of smaller absolute value are considered greater (e.g., $-5 > -6$). We posed integer-comparison tasks to 40 students each in Grades 2, 4, and 7, as well as to 11th graders on a successful mathematics track. We coded for correctness and for students' justifications, which we categorized in terms of 3 ways of reasoning: magnitude-based, order-based, and developmental/other. The 7th graders used order-based reasoning more often than did the younger students, and it more often led to correct answers; however, the college-track 11th graders, who responded correctly to almost every problem, used a more balanced distribution of order- and magnitude-based reasoning. We present a framework for students' ways of reasoning about integer comparisons, report performance trends, rank integer-comparison tasks by relative difficulty, and discuss implications for integer instruction.

1. Introduction

Integers are an important and challenging topic in the transition from arithmetic to algebra (Peled & Carraher, 2007). Children first learn about whole numbers, which are rather intuitive because they relate to counting and quantifying sets of items in the world. Against this backdrop, the notion of a negative number—often described as being "less than zero"—requires some suspension of disbelief. In many everyday contexts, such as numbers of people, toys, cookies, and so on or measures such as length or area, the idea of a number less than nothing seems absurd. How then do students make sense of integers? In particular, what kinds of justifications do they offer for their judgments that one integer is greater than or less than another?

We interviewed 40 students each in Grades 2, 4, 7, and 11 and asked them to compare pairs of integers, such as −7 and 3. In this paper, we report on the justifications that students offered for such comparisons; we categorize the justifications as belonging to broader ways of reasoning about integers. We compare and contrast students' reasoning by looking both across problems and across grade levels. In this way, we answer how students reason about different cases of integer comparisons

☆ This material is based upon work supported by the National Science Foundation 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.
* Corresponding author.
E-mail address: iwhitacre@fsu.edu (I. Whitacre).

http://dx.doi.org/10.1016/j.jmathb.2016.11.001
0732-3123/

and identify trends in the reasoning of students with different levels of familiarity with negative numbers. We conclude with implications for integer instruction that are informed by trends in students' reasoning.

In reviewing the literature, we found a lack of literature that systematically sampled K-12 students at different grade levels and documented their explicit reasoning about integer comparisons. We present a framework of students' ways of reasoning and associated justifications for integer comparisons. We also report compelling trends in students' reasoning both within and between groups. Given the counterintuitive nature of negative numbers, understanding students' thinking about them is particularly important for supporting student learning. The findings reported here advance the field's understanding of students' thinking about integers. Such research contributes to the efforts of the mathematics education research community to support instruction that enables students to successfully transition from arithmetic to algebra (e.g., Moses & Cobb, 2002; Peled & Carraher, 2007).

2. Theoretical perspective

We approach this study from a children's mathematical-thinking perspective (Steffe, 1991). We regard children's mathematical thinking as being different from that of adults in interesting and important ways. We take seriously the nature of children's mathematics, whether or not it is correct from an expert perspective. We believe that seeing mathematics through children's eyes is important for better understanding the sense that they make. This perspective is based on constructivist principles that children have existing knowledge and experiences they bring with them into the classroom and upon which they continue to build (e.g., Carpenter, Fennema, Franke, Levi, & Empson, 1999; Fuson, Smith, & Lo Cicero, 1997; Steffe & Olive, 2010; Steffe, 1991, 2002, 2004). We take this view because the ultimate goal of our research is to find ways to better support children's learning of mathematics (Carpenter et al., 1999; Carpenter, Franke, & Levi, 2003; Empson & Levi, 2011).

3. Background

In the early elementary grades, students become acquainted with whole numbers and the basic operations involving these. While they progress in their mathematical educations, students encounter different kinds of numbers. Whole-number arithmetic is relatively intuitive for children because they can reason about it in ways that are grounded in real-world contexts (e.g., Carpenter et al., 1999). (Nonnegative) fractions, decimals, and percentages present additional challenges, because children are asked to reason about numbers with values between the whole numbers, including between 0 and 1. Furthermore, these numbers are represented in many ways, and children must learn to relate the various representations. At the same time, nonnegative rational numbers can still be related to amounts in real-world contexts (e.g., 3/4 of an apple pie). As they do with whole numbers, children can draw on their intuitions and physical experiences to make sense of fractions and to represent them in multiple ways (e.g., Empson & Levi, 2011). With the introduction of negative numbers, new challenges arise.

In the United States, instruction on integer arithmetic is typically concentrated in middle school (National Governors Association Center for Best Practices [NGA] & Council of Chief State School Officers [CCSSO], 2010; Whitacre et al., 2011). By that time, students' comfort levels with nonnegative numbers may work against them (Bruno & Martinón, 1999). They are asked to expand their mathematical worlds to include negatives, as well as to reconceive of familiar numbers as positive (Whitacre et al., 2016). Previous generalizations cease to be true, and the bounds of mathematical reality are challenged. For example, children who think of −7 as representing 7 of something are asked to see −7 as *less than* 3, but how can 7 of something be less than 3 of something? Understandably, the introduction of integers involves notions that are counterintuitive for children (e.g., Vlassis, 2004).

3.1. Magnitude and order

When considering the pedagogical challenges of supporting her students' understanding of negative numbers, Ball (1993) wrote,

> Any number has two components: magnitude and direction; from a pedagogical point of view, this seems to become particularly significant when the students' domain is stretched to include negative numbers. A focus on the magnitude component leads to a focus on absolute value. This component emerges prominently in many everyday uses of negative numbers (e.g., debt, temperature). Thus, comparing magnitudes becomes complicated. There is a sense in which −5 is more than −1 and equal to 5, even though, conventionally, the "right" answer is that −5 is less than both −1 and 5. This interpretation arises from perceiving −5 and 5 as both five units away from zero and −5 as more units away from zero than −1. Simultaneously understanding that −5 is, in one sense, more than −1 and, in another sense, less than −1 is at the heart of understanding negative numbers. (p. 379).

Historically, mathematicians struggled to resolve the contradiction between magnitude and order in the interest of consistency (Gallardo, 2002; Henley, 1999). This dilemma was resolved by the adoption of a more abstract notion of number and a convention that privileges order over magnitude. In mathematics today, the symbols < and > and the corresponding terms *less than* and *greater than* refer to comparisons of order rather than magnitude. In other words, the statements "2 > −10" and equivalently "2 is greater than −10" mean that 2 is to the right of −10 on the number line. Technically speaking, these statements have everything to do with order and nothing to do with magnitude.

Throughout this paper, we make use of the distinction between magnitude and order. Consider first the comparison of a pair of nonnegative integers, such as 5 and 6. Thinking in terms of magnitude, six things are more than five things (e.g., things could be cars, or marbles, or any countable entity). In other words, *comparing in terms of magnitudes* means comparing the cardinalities of sets. Thinking in terms of order, by contrast, 6 is to the right of 5 on the number line, or 6 comes after 5 when counting. Thus, *comparing in terms of order* means comparing based on the established sequence of number words, symbols, or locations. Clearly, magnitude and order are distinct foci; yet, for nonnegative numbers, they are consistent. Two people may reason differently when comparing 5 and 6, yet they will agree that 6 is greater than 5, and the difference in their reasoning may go unnoticed. Likewise, one person may at times reason in terms of magnitude and at other times in terms of order, or in both ways in the same instance, without any conflict.

When one compares negative integers, the distinction between magnitude and order becomes much more interesting because now the judgments conflict. A student who thinks of −6 as 6 (negative) things may reasonably conclude that this number is more than 5 (negative) things. (Indeed, |−6| is greater than |−5|.) By contrast, a student who reasons that −5 is further to the right on the number line, or is closer to 0, may reasonably conclude that −5 is greater than −6. (This order-based comparison is conventionally considered to be the correct one.).

3.2. Previous research on integer comparisons

We summarize literature from two research traditions with different approaches to examining how people compare integers. First, we review mathematics education research that is focused on students' reasoning about integer comparisons or related tasks during interviews or instructional activities. Second, we review psychological experiments concerned with testing hypotheses about mental processes or mental representations that might explain patterns in people's responses to integer-comparison tasks. We consider what both of these traditions show us about how students compare integers, and we identify questions that have previously gone unanswered in the literature. We frame our review of the literature using the distinction between magnitude-based and order-based reasoning.

3.2.1. Mathematics education research concerning integer comparisons

Peled and Carraher (2007) argued that signed numbers should be introduced in the early grades to support students' learning of algebraic concepts. Indeed, some studies have involved elementary students in reasoning about integers or integer-related tasks before the topic is traditionally introduced (e.g., Bishop, Lamb, Philipp, Whitacre, & Schappelle, 2014; Bofferding, 2014; Hativa & Cohen, 1995). The focus of studies summarized below is on students' using order-based reasoning to extend their number knowledge to include negative integers. Because few articles are focused on integer comparisons, we include studies concerning integer arithmetic more broadly.

3.2.1.1. Affordances of order-based reasoning. A group of articles illustrate the affordances of order-based reasoning. Wilcox (2008), in analyzing the thinking of a first grader playing a number-line game, highlighted the point that reasoning about numbers as locations on a number line affords the possibility of considering numbers to the left of zero. Wilcox argued that, by contrast, magnitude-based reasoning might introduce a barrier to the learning of integers because "there is nothing less than zero" (p. 204). Thus, order-based reasoning allows for numbers to the left of zero, because the meaning of *number* is not tied to magnitude.

Bishop, Lamb, Philipp, Schappelle, and Whitacre (2011) described how some first graders extended their thinking beyond whole numbers to reason about numbers less than zero. Students playing a number-line game invented names for locations to the left of zero. A student who was already familiar with negative numbers extended his whole-number counting strategies into the negatives (e.g., to solve $3 - 5 = \square$ by counting backward). Another student used the idea of motion on the number line to support reasoning about integer operations. In these examples, order-based reasoning was particularly conducive to productive strategies. By contrast, those students who relied on magnitude-based reasoning were perplexed by counterintuitive problems wherein, from a magnitude-based perspective, one must subtract more than one has at the beginning. For example, one student's response to $3 - 5 = \square$ was, "You can't do it, because you can't take five from three" (p. 354).

Bishop et al. (2014) documented how a second grader called Violet built upon order-based reasoning to solve unfamiliar integer-arithmetic problems. Violet initially recognized negative numbers as locations on a number line but reasoned about addition and subtraction in terms of generalizations based on her experiences with whole numbers, such as the idea that addition makes larger. She extended from this conception of number to a conception that included the idea that negatives do the opposite of positives. In terms of movement on a number line, addition (or subtraction) of a negative number would move in the direction opposite from addition (or subtraction) of a positive number.

Bofferding (2014) investigated first graders' mental models of integers, focusing on how they ordered integers and compared their values. She found a hierarchical progression from students who ignored negative signs to those who focused on absolute value to those who compared integers on the basis of order. At the most advanced level, students have what Bofferding called an *integer conception,* which entails reasoning about integer comparisons in terms of order, rather than magnitude. She found that 71% of first graders initially ignored negative integers or treated them as whole numbers (e.g., treating −5 as 5). Thus, most of the students were essentially unaware of negative numbers. Bofferding found that 23% of first graders had some knowledge of negative numbers and could at least compare negative numbers to positive numbers. Only 11% exhibited integer conceptions and were able to consistently order and compare negative numbers across several types of ordering and comparison tasks.

Bofferding compared three instructional interventions aimed at developing first graders' conceptions of integers. She found that students in all three groups improved. The biggest improvements were among the group that played games involving ordering integers on number lines and participated in activities focused on comparing mathematical expressions such as −4 – 5 and −4 – −5 to distinguish subtraction and negative signs. Bofferding's findings indicate potential for students in the primary grades to engage productively with integer instruction and to advance their conceptions. We return to this point in the *Discussion* section.

3.2.1.2. Affordances of magnitude-based reasoning. The above articles illustrated the affordances of reasoning about integers in an order-based way. The following articles emphasize the affordances of magnitude-based reasoning.

Whitacre et al. (2012) used an opposite-magnitudes context to investigate K-5 students' integer-related reasoning. In a context of happy and sad thoughts that cancel each other, the authors asked students to evaluate the happiness or sadness of various hypothetical days. They reported three distinct ways of reasoning observed in students' responses, documenting increasing sophistication and formality across the grades. In this context, magnitude-based reasoning was quite productive, provided that it took into account the positive (happy) or negative (sad) quality of those magnitudes (numbers of days). Thus, contextual meanings associated with negatives can clarify that negative numbers that are larger in absolute value should be considered lesser rather than greater, because they are less desirable or have a "negative connotation" (Peled, 1991).

Linchevski and Williams (1999) reported on the reasoning of sixth graders in two teaching experiments. The students had not previously experienced instruction involving negative numbers, and the teaching experiments introduced them to signed numbers through the use of one of two contexts: a disco game or a set of dice games. The authors demonstrated how students both developed meanings for integers as representing a state or net effect and reinvented rules for integer arithmetic on the basis of the contextual meanings of addition and subtraction, leveraging magnitude-based reasoning in both contexts. In this way of reasoning, ideas of zero pairs ($1 + -1 = 0$) and equivalent sums (e.g., $-5 + 3 = -6 + 4$), relevant to the learning of algebra, are emphasized (Peled & Carraher, 2007).

3.2.1.3. Integrating order and magnitude. Both order-based and magnitude-based reasoning can support students' solutions to tasks involving integers. The following two articles illustrated how these ways of reasoning may be integrated. Stephan and Akyuz (2012) documented the evolution of collective activity in a seventh-grade class during a unit on integer arithmetic. Instruction during the unit was focused on the context of net worth. Students calculated and compared people's net worths, and they solved problems involving changes in net worth. In particular, they used zero as a reference point to compare negative amounts. For example, students argued that a net worth of −$20,000 should be placed higher (on a vertical number line) than a net worth of −$22,000 "because −20,000 is being closer to out of debt than −22,000" (p. 453). Thus, in activity involving the context of net worth, students brought both magnitude-based and order-based reasoning to bear when comparing negative numbers.

Bishop et al. (2014) looked to both the history of mathematics and interviews with elementary school students to identify obstacles and affordances for integer reasoning. They found that magnitude-based reasoning was a barrier to the acceptance of negative numbers historically, because of the difficulty of conceiving of numbers less than nothing. Acceptance of negative numbers came with a shift to more abstract and formal views of number. Bishop and colleagues documented how children face some of the same challenges as mathematicians did historically when they are confronted with negative numbers or with problems that seem counterintuitive because of their previous experiences (e.g., $5 + \square = 3$). They found that children who reasoned successfully about integers overcame obstacles in three ways: by leveraging order, by using formal (logical) reasoning, and by reasoning in terms of magnitude. Despite the appearance of magnitude-based reasoning as an obstacle to the acceptance of negative numbers, children used magnitude productively, such as to solve $-5 - -3 = \square$ by analogy to the behavior of "normal numbers" in which $5 - 3 = 2$ (p. 51). Bishop et al. thus documented affordances of both order-based and magnitude-based reasoning in children's attempts to make sense of integers and integer arithmetic.

Looking across the studies summarized above, we see an interest in how students conceive of notions of *greater than* and *less than* when negative numbers are involved, as well as some attempts to encourage the development of order-based, magnitude-based reasoning, or both. In certain contexts, such as happy and sad thoughts, reasoning about integers in terms of magnitude can be quite productive for students. Magnitude contexts can also be connected with order-based reasoning, as in the use of a vertical number line to represent net worth (Stephan & Akyuz, 2012).

In the above studies, the affordances and constraints of order-based and magnitude-based reasoning have been implicitly or explicitly investigated; however, this literature leaves many questions unanswered. In terms of documenting how students reason about integer comparisons, we see the following limitations to the previous literature: (a) most investigations have focused on integer arithmetic, rather than integer comparisons; (b) most studies have focused on students at a single grade level; (c) these studies have not produced results that allow for comparisons between grade levels; and (d) apart from the work of Bishop et al. (2014), these studies were conducted with small groups of students or with students at a single school.

Our work is motivated by an interest in how students reason about integer comparisons. We addressed the above limitations by (a) conducting a systematic investigation focused on students' reasoning about integer comparisons, (b) collecting data from students at multiple grade levels, (c) using the same tasks and coding scheme with each group of students, so that the results may be compared, and (d) interviewing a range of students from different schools in a deliberate attempt to collect a representative sample of data on students' reasoning.

3.2.2. Psychological experiments concerning integer comparisons

Researchers in experimental psychology and mathematical cognition have studied how people respond to integer-comparison tasks. In this research tradition experiments are laboratory style, often focused on response time, designed to test hypotheses regarding the mental processes and mental representations that people may use when comparing integers (see Krajcsi & Igács, 2010; Table 1, for a concise summary of the methods and results of several such studies). Two competing hypotheses feature prominently in this literature, and both suppose the use of a mental number line. According to the *phylogenetic*[1] hypothesis, on the one hand, the mental number line that people use when comparing integers consists only of natural numbers. When asked to make comparisons involving negatives, people use this mental natural-number line along with a set of learned rules for operating with negatives. For example, to compare −5 with −6, a person would consult his or her mental natural-number line to know that 5 < 6 and then would reverse this inequality for negatives to answer that −5 > −6. According to the *ontogenetic* hypothesis, on the other hand, the mental number line is extended (over time and while people learn) to include negative numbers. Then this mental integer-number line is used to make all comparisons involving integers. So, a person would compare −5 with −6 on the basis of the relative positions of these numbers on the mental number line: −5 is to the right of −6, so −5 > −6. Various researchers have tested these and related hypotheses for the mental processing involved in comparing integers (e.g., Fischer, 2003; Ganor-Stern & Tzelgov, 2008; Krajcsi & Igács, 2010; Parnes, Berger, & Tzelgov, 2012; Varma & Schwartz, 2011).

Researchers have reported *magnitude effects* in which the absolute values of the given integers in a comparison task influence response time (Fischer, 2003; Gullick & Wolford, 2013). Magnitude effects are well known for natural numbers. For example, people take longer to compare numbers that are close together (e.g., 1 cf. 4) versus numbers that are farther apart (e.g., 1 cf. 9). Gullick and Wolford (2013) focused on positive–positive and negative–negative comparisons. In cases of *mixed* comparisons (involving a positive and a negative number), the situation is more complicated. Because of conflicting study results, whether or how the distance between the numbers affects response time for mixed comparisons is unclear (Varma & Schwartz, 2011).

Researchers have also reported *judgment effects,* the most notable of which is the *Spatial-Numerical Association of Response Codes (SNARC) effect*, which has to do with associating small or negative numbers with left space and large or positive numbers with right space. For example, some researchers asked subjects to press either the left-most or right-most key on a keyboard to indicate their responses to a comparison task. When comparing natural numbers, people responded faster on the left side when comparing small numbers, and they responded faster on the right side when comparing larger numbers (e.g., Fischer, 2003). The SNARC effect is generally regarded as evidence of the use of a "mental number line," which may or may not include negative numbers (e.g., Gullick & Wolford, 2012; Huber, Cornelsen, Korbinian, & Nuerk, 2015; Parnes et al., 2012).

The present study was conducted in a different research tradition than that of the studies described above. One major difference is that we are not interested in response time. We presented integer-comparison tasks in a relatively relaxed interview setting (without time pressure) and asked students to explain their reasoning about each comparison task. We used students' answers (the numbers that they selected as being greater) together with video-recorded evidence of their responses to categorize their reasoning and to code their specific justifications. We are not seeking a one-size-fits-all theoretical model to explain how people compare integers. Indeed, we see no need to assume that a single mental process or mental representation should explain people's reasoning. On the contrary, we are especially interested in differences in students' justifications and ways of reasoning, both between grades and within grades.

We focus on the reasoning of K-12 students and investigate trends by grade level, whereas most of the studies in the above research tradition involved college students. College students are generally quite adept at integer-comparison tasks. As a result, error rates tend to very low. We are interested in the reasoning of students both before and after integer instruction because we value student-centered instruction, which builds upon students' ideas and approaches. To design and enact such instruction, educators need to know how students may reason about the relevant tasks on the basis of their prior knowledge. Thus, we are interested (among other aspects) in the reasoning of students who give incorrect answers.

We found few articles in the psychological literature that reported on the responses of K-12 students to comparison tasks involving negative integers. Gullick and Wolford (2013) studied the responses of fifth and seventh graders to integer comparisons between −20 and 20, focusing on positive–positive and negative–negative comparisons. They reported, "Between subjects, there was no significant effect of age . . . indicating that fifth-grader (mean = 84.8%) and seventh-grader (mean = 88.3%) performance was similar on the task" (p. 6). This result is surprising, given that the researchers were ostensibly comparing a preinstruction and a postinstruction group. One might reasonably expect the preinstruction group to significantly underperform the postinstruction group. Evidently, even the preinstruction group was familiar with negative integers, or perhaps some feature of the task design supported them in responding correctly.

Similarly, Varma and Schwartz (2011) posed integer-comparison tasks to 36 U.S. sixth graders who were in an accelerated mathematics class. They reported an error rate of less than 4%. This very low rate may have been due to (a) the researchers'

[1] Authors have used a variety of names and descriptions to refer to these hypotheses.

procedure of providing real-time feedback to participants on the correctness of their responses after each comparison task or (b) the choice of student population.

Although the two studies above involved K-12 students—the population of interest in our work—the results of those studies and the others above leave many unanswered questions: How do students perform on integer-comparison tasks when they have little or no familiarity with signed numbers? Are there differences in the reasoning of students within or between student groups? How do students themselves describe their reasoning about integer-comparison tasks? What justifications do they offer for their claims that one integer is greater than another, and what ways of reasoning do these indicate?

4. Method

Building on the literature reviewed and the ideas discussed above, we addressed the following research questions: How do students in Grades 2, 4, 7, and 11 reason about integer comparisons? More specifically, (a) How do students perform when responding to different types of integer comparisons? (b) What kinds of justifications to do they offer for their judgments, and how frequent are these justifications? (c) How do ways of reasoning relate to performance? Below we describe the methodological details of this study.

4.1. Setting and participants

We conducted clinical interviews with students in Grades 2, 4, 7, and 11 during the spring of 2011. We selected a range of grade levels to investigate trends in integer reasoning to be found at these different levels. At the time and location of our study, most formal integer instruction was concentrated in Grades 5–7. The 2nd and 4th graders had not received formal instruction on integers and, therefore, would be a source of information about children's intuitive and informal ways of reasoning. We interviewed 7th graders in May, near the end of their school year, so that they would already have experienced instruction focused on integers in Grades 5–7. We also interviewed 11th graders who were enrolled in precalculus or calculus and, thus, were on a successful track in school mathematics.

The 160 students in Grades 2, 4, 7, and 11 were from 11 Southern California public schools (3 elementary schools, 3 middle schools, 1 K–8 school, and 4 high schools) with a range of standardized-test scores as indicated by each school's Academic Performance Index (API). We purposefully selected schools to ensure the inclusion of those with varied demographics. At each participating school, all students in the classrooms of two teachers at each targeted grade level were invited to participate. Students were then selected randomly from among all who returned signed consent forms. To reach the total of 40 students per grade level, we selected 9–11 students per grade level from each of the four schools associated with each grade level. In our analysis, we do not attempt to relate specific instructional experiences to students' conceptions. Rather, we sought a representative sample of students to describe the range of conceptions observed and the frequencies of these conceptions.

We categorized students into one of four groups, named *college-track 11th graders, 7th graders*, *2nd and 4th graders with negatives*, and *2nd and 4th graders without negatives*. The 7th and 11th graders are simply grouped by grade level because all in each group had received formal instruction on integers and integer operations. The 2nd and 4th graders are grouped on the basis of whether we saw evidence that they had some knowledge of negative numbers in their responses to tasks posed in the beginning part of the interview. In particular, these groups were determined primarily on the basis of responses to the following two tasks.

1 *Name a small number.* After the child responded, the interviewer asked, *Can you name a smaller number?* If the child responded, “One,” the interviewer asked, *What if I gave that away? What number would you have then?* If the child responded, “Zero,” the interviewer asked, *Is there a number smaller than zero?*
2 *Can you count backward, starting at 5?* If the child stopped at 0 or 1, the interviewer asked, *Can you keep counting back?*

Our sample of 40 students each in Grades 2 and 4 resulted in 41 students in the group of 2nd and 4th graders without negatives and 39 students in the group of 2nd and 4th graders with negatives. For the purposes of analysis, we categorized students in this way because we believe that their degree of familiarity with negative numbers is the factor most relevant in this study.

4.2. Measures and methods of analysis

In this paper, we focus on students' responses to the following integer-comparison tasks:

3	7
−7	3
−5	−6
−9	0
−5	−100

We focused our analyses on the comparison tasks involving one or two negative integers. We included the first comparison (3 cf. 7) as a control. The pairs −5 compare −6 and −5 compare −100 were chosen to test whether students performed differently depending on the distance between two negative numbers. The other number combinations were chosen to represent distinct cases (e.g., negative cf. zero and negative cf. positive). Apart from −5 compare −100, we used single-digit numbers to avoid conceptual complexities associated with place value in multidigit numbers.

4.2.1. Prompts

The tasks were printed on a sheet of paper, and they were posed to students one at a time. The interviewer verbally provided the directions: *For each pair of numbers circle the larger, write = if they are equal, or write? if there is not enough information to tell which one is larger.* The interviewer then asked, *Can you read each pair of numbers for me so that when we watch you on the video we know which problem you are solving?* After the student marked a response, the interviewer asked, "How did you think about that?" and then posed additional follow-up questions if necessary to understand how the student determined each comparison.

4.2.2. Students' justifications for integer comparisons

Typical integer-comparison tasks (like those that we used in this study) do not require computation; they simply require students to decide upon and make a claim about one number relative to another—greater than, less than, or equal to. The nature of these tasks contrasts with that of computational items, which are more open-ended in terms of potential answers and which require some computational work. Because of the nature of comparison tasks, we speak in terms of students' justifications for integer comparisons as opposed to strategies for making comparisons. We are interested in how students' justify their comparison claims because we regard these justifications as a window into students' underlying ways of reasoning about integer comparisons. For example, Student A claims that −6 is greater than −5 and offers the justification that 6 is more than 5, along with the example that 6 fingers is more than 5 fingers. Such a justification reveals that Student A is reasoning about the integer comparison in terms of number magnitude, rather than order. By contrast, Student B argues that −5 is greater than −6 because −5 is closer to 0. Student B is reasoning about this comparison in terms of order, as opposed to magnitude. We are interested in both the particular justifications that students offer and the underlying ways of reasoning that these reveal.

4.2.3. Coding scheme

Each student response was coded for correctness and justification for that answer, which belonged to a broader way of reasoning. Justification codes were generated through open coding of a subset of the data. We sought fine-grained codes that distinguished nuances of students' reasoning. These codes were later grouped according to three broader ways of reasoning: order-based reasoning, magnitude-based reasoning, and developmental/other reasoning. The development of the coding scheme was both grounded in the data that were initially analyzed and informed by our review of the literature.

We used the following scheme to code the responses of all participants. We coded directly from the video recordings of the interviews and referred to student work collected during the interviews as needed. Throughout the coding process, 20% of the interviews were randomly assigned to be double coded. Because the coding assignments were made by a project team member who did not participate in coding, coders were blind to which interviews would be double coded. Using these double-coded data, we checked reliability by comparing coding decisions at the fine-grained level of justifications, rather than the broader level of ways of reasoning. Coders agreed on 89% of coding decisions for students' justifications of their integer comparisons. Discrepancies were resolved through discussion.

In particular, we refined the coding by revisiting any disagreements that seemed to reflect differences in coders' conceptualizations of the ways of reasoning. We took a particular interest in responses that were coded as order-based by one coder and magnitude-based by another. In these instances, coders reviewed and recoded the data and then argued for their interpretations. These discussions led to clarification of operational distinctions between the ways of reasoning, which are reflected in the definitions that we present below.

4.2.3.1. Order-based reasoning about integer comparisons. Order-based reasoning involves reasoning about integers in terms of their places in a counting sequence or their locations on a number line. Students using order-based reasoning focused on which number came before or after the other number when counting, or which was further to the right, further to the left, or closer to 0 on the number line. In particular, the following order-based justifications were identified.

4.2.3.1.1. Negative integers are less than positive integers or zero. Students who gave this justification compared negative integers to nonnegative integers on the basis of the general principle that negative numbers are always less than positive numbers or zero or that positive numbers are always larger than negative numbers or zero. For example, Alana, a 7th grader, compared −7 and 3 by circling 3, and then stating, "Well, the negative side is always going to be less than positive." In this example, Alana appeared to reference the number line, by indicating that the negative "side" is less than the positive. Note that many children are not familiar with the term *positive numbers* and so often use alternative language, such as "regular" numbers to refer to the set of natural numbers. Students in every participant group used a version of this justification. This justification corresponds to the reasoning of students who are at the level of the *Synthetic mental model* or above (*Transition*

II or *Formal*) in Bofferding's (2014) scheme. It is also consistent with the phylogenetic hypothesis, in which people apply a rule for comparing negative and positive integers (Varma & Schwartz, 2011).

4.2.3.1.2. Closer to 0 or closer to positive numbers. Students who gave this justification compared two negative integers by reasoning that the number closer to 0 (or to the positive integers) would be greater. Note that a closer-to-zero justification is valid for comparisons of two negative integers, but it would consistently lead to incorrect conclusions if applied to comparisons of a negative integer with a positive integer. In identifying this justification, we are not saying that students who used it believe that it holds for all integers. On the contrary, students tended to use it appropriately. For example, Caden, a 4th grader, selected −5 as being greater than −6, saying, "I bet this stumps a lot of kids. It is actually negative 5, because negative 5 is closer to 0, and negative 6 is one more away from 0." Caden's use of 0 as a reference point involves taking into account numbers that are closer to and further from that reference point (in this case, 0), and so this justification is based on the ordered nature of the number system. Stephan and Akyuz (2012) described students comparing negative net worths by referencing distances from zero. Bofferding (2014) considered students who consistently reason in this way to be using the *Formal mental model* that she referred to as *Integer*. This justification is consistent with the ontogenetic hypothesis, in which people extend their mental number lines to include negative number-locations (Fischer, 2003).

4.2.3.1.3. Greater further right. Students who gave this justification compared two integers by reasoning that the number further to the right was greater than, or that the number further to the left was less than, its counterpart. Students who reasoned this way sometimes referred explicitly to a number line or to the counting sequence, gestured to indicate direction, or both. For example, a 7th grader named Kylee selected −5 as greater than −6 and explained, "I'd say negative 5 is larger, because it's like one number to the right than negative 6." She indicated that her determination was made on the basis of −5 being "one number *to the right*" of −6. Note that this justification is the most generalizable in that it applies to all comparisons of two integers. Regardless of their specific locations on a number line, numbers further to the right are considered to be greater. Thus, this justification is related to *Closer to 0 or to positive integers*, except that it applies consistently across the number line. Again, students who consistently reason in this way are using the *Formal mental model* that Bofferding (2014) referred to as *Integer*. This justification is especially consistent with the ontogenetic hypothesis, because it does not require a special rule for dealing with negative numbers (Fischer, 2003), but instead indicates reference to a unified integer number line.

4.2.3.1.4. Greater further left. Students who provided this justification compared two negative integers by reasoning that the number further to the left on the number line was greater. This reasoning looked similar to the justification *greater further right*, except that it led to opposite conclusions. Some students who used *greater further left* appeared to reason about the negative side of the number line as a mirror image of the positive side, so that negative numbers further to the left are greater, just as positive numbers further to the right are greater. For example, a 2nd grader named Bradley selected −6 as greater than −5. He then shared his thinking:

> I think the larger number is negative 6, because negative 5 is right here [pointing to −5 on number line] and negative 6 is right here [pointing to −6 on number line], and whatever one is bigger is on the left side [motions to the left] in the negatives. All the numbers that are bigger, like, in negatives, it [the larger number] is going left.

In this example, Bradley articulated that an aspect of his justification is that he is comparing numbers "in the negatives," and when comparing negative numbers, one knows that larger numbers are further left on the number line. His reference to the number line and his sweeping motion toward the left provide evidence that this is an order-based justification.[2] This justification is consistent with Bofferding's (2014) *Synthetic mental model* in which students consistently make incorrect comparisons of pairs of negative integers (although they consistently make mixed comparisons correctly). The relationship of this justification to the psychological literature is less clear, because that literature is focused on response times for correct responses; this justification, by contrast, consistently yields incorrect answers. The justification could be a case of a special rule for dealing with negative integers, together with knowledge of natural numbers (i.e., phylogenetic), or it could be a case of a mental number line extended to include negatives but with properties that do not align with mathematical conventions (i.e., ontogenetic).

4.2.3.2. Magnitude-based reasoning about integer comparisons. Magnitude-based reasoning involves reasoning about numbers as representing amounts or cardinalities. Such reasoning may or may not involve analogies to specific objects or contexts. Students who gave magnitude-based justifications compared magnitudes (or absolute values) of numbers. Consistent with this way of reasoning, students reasoned about zero as nothing, some also considered negative numbers to be nothing, and some related integers to specific contexts. In particular, the following four magnitude-based justifications were identified:

4.2.3.2.1. Compares magnitudes. Students who gave this justification compared two negative integers by comparing them to positive integers or by thinking of negative integers as being more or less negative than another negative number. For example, some students claimed that −6 was larger than −5 by explicitly comparing that pair of numbers to 6 and 5. Other students reasoned that −6 was *more negative* than −5. For this rationale, students might use that justification to support choosing either −6 or −5 as being greater. For example, 7th grader Ida selected −5 as greater than −100, saying, "I think it's

[2] Exactly what students mean by terms such as *larger* or *greater* is often unclear. Responses such as Bradley's may include elements of both order-based and magnitude-based reasoning; however, given his explicit reference to the number line, we consider this justification to be order-based.

negative 5 because the bigger the number in negatives, the smaller it actually is." By contrast, 2nd grader Betty (who was familiar with negative numbers) claimed that −100 was greater than −5. She said, "That's a negative 100. Because 100 is bigger than 5." Both Ida and Betty compared the magnitudes of −5 and −100, but they drew opposite conclusions.

This tendency to focus on magnitudes in comparing negative integers has been identified by mathematics education researchers including Ball (1993) and Bofferding (2014). This justification is consistent with the phylogenetic hypothesis in which students rely on their knowledge of natural numbers, together with rules for dealing with negative numbers. On the one hand, if such rules are not known or are incorrectly remembered, students might claim that negative integers of greater absolute value are greater. On the other hand, this justification might also be regarded as consistent with the ontogenetic hypothesis, given again that the students have not yet learned to deal correctly with negatives and have not extended their mental number lines.

4.2.3.2.2. Zero is nothing. Students who gave this justification compared a nonzero integer to zero by reasoning that zero is nothing. For example, Arnold, a fourth-grade student, selected −9 as greater than 0, explaining, "Zero doesn't have a meaning and negative 9 does. It's [−9] a part of a whole, and zero isn't." This example is an instance of a magnitude-based view of number in that the negative number is deemed larger because zero has no magnitude (in Arnold's words, "Zero doesn't have a meaning"), whereas the negative integer *has* magnitude, as inferred by its selection as the larger of the two numbers. This justification is consistent with Bofferding's (2014) Initial mental models of integers in which 0 is treated as the smallest number, either because negative signs are ignored or numbers are compared according to their absolute values. As with *compares magnitudes*, this justification may be regarded as consistent with both the phylogenetic and ontogenetic hypotheses.

4.2.3.2.3. Negative numbers are nothing. Students who gave this justification indicated that negative numbers had no value. A child might have determined that all positive integers are larger than negative integers because negative integers have no magnitude (they are "nothing"). Some students explicitly treated all negative numbers as being equal to 0. For example, 4th-grader Cara compared −9 to 0 by placing an equal sign between the two numbers and saying, "I think it's equal. Because, the negative sign is like zero, cause it's like taking away. So, it's like the 9 is a zero and this is a zero (points to zero), so zero and zero is equal." The fact that students may treat negative numbers as zero was reported by Bofferding (2014) and corresponds especially to the *Transition I mental model* called *Conflicted value* in her scheme. Again, the psychological literature seems to have little to say about justifications such as this one because that research has been focused on competent performance.

4.2.3.2.4. Converts to context. Students who gave this category of justification contextualized the numbers. For most contexts, students drew on a magnitude-based view of numbers, such as dollars owed. For example, some students argued that −5 is greater than −6 because it would be better to owe less money. However, others were more creative. One 7th-grade student, Fay, compared −7 and 3 by relating the numbers to Oreo cookies. She said, "Three is bigger because this [pointing to −7] is negative. Would you rather have, like, [pointing to −7] crumbled Oreos, all nasty old, moldy, or 3 good Oreos? [answering her own question] The good Oreos." In her comparison of −5 and −6, she continued using this context. Circling −5, Fay stated, "I would rather have negative 5 crumbly old Oreos rather than 6." Fay used a magnitude-based context for her comparisons, by comparing positive and negative numbers with good and "nasty, moldy" Oreos, respectively.

Many authors have addressed students' reasoning about contexts involving directed magnitudes, or opposite magnitudes, and in some cases relating these to integers (e.g., Bofferding, 2014; Linchesvski & Williams, 1999; Stephan & Akyuz, 2012; Whitacre et al., 2012). The justification *converts to context* is different from those cases in that it involves the student's invoking a context to reason about a naked-number task. Furthermore, this reasoning did not take place in a mathematics class or teaching experiment during an integer unit, and many of the contexts that students used were nonstandard (i.e., not identified in the CCSSM or typically used in textbooks (NGA & CCSSO, 2010; Whitacre et al., 2012)). How such justifications relate to the results of psychological studies is unclear; in those studies participants respond under time pressure and do not give verbal explanations. Researchers analyzing such studies tend to assume use of a mental number line and do not seem to allow for or take an interest in reasoning such as *converts to context.*

4.2.3.3. Developmental/Other reasoning about integer comparisons. *Developmental/other* reasoning is typical of students who are unfamiliar with negatives. Justifications in this category were based on students' knowledge of whole numbers and operations. Some students simply ignored minus signs (e.g., treating −7 as 7). Others treated minus signs as indicating subtraction. Some simply stated that there were no such numbers. We identified the following two justifications associated with this way of reasoning:

4.2.3.3.1. Ignores sign(s). Students who gave this type of justification compared a pair of integers by treating them as natural numbers. If minus signs were present, the student simply ignored these. Typically, students whose responses received this code did not read the minus sign (e.g., they read −7 and −6 as simply "seven" and "six"). This category also includes the response of a child who read −7 as "I don't know, 7" or "line 7" but reasoned about the comparison as if the minus sign had no effect. Tim, a 4th grader, selected −7 as greater than 3, stating, "Because 7 is a bigger number than 3." This justification corresponds to Bofferding's (2014) *Initial: Whole number* mental model of integers in which "negative integers [are] ignored or treated as positive integers" (p. 222). As with other forms of incorrect justifications, relating these to the psychological literature is difficult.

4.2.3.3.2. Treats signs as operations. Students who gave this type of justification treated the minus sign as an operation rather than as the sign of the number. For example, 4th-grader Kim selected 3 as greater than −7, explaining, "'Cause minus

7 [referring to −7], you are taking away 7, so there would be nothing left, so the 3 would be higher.” The conversation continued.

Interviewer: Oh, okay. So, there would be nothing left? [Interviewer points to −7]

Kim: Yeah, nothing but a zero. So, if it was a zero, then the 3 would be higher.

In this example, Kim treated the minus sign as an operation, chose to subtract 7 from 7, and compared that difference of 0 to the 3 to determine that 3 was the larger number. Although Kim’s answer using this reasoning was correct, the response was assigned the code “Correct answer for wrong reason,” because −7 is not equal to 0. This justification also corresponds to the *Transition I mental model* of *Conflicted value* in Bofferding’s (2014) scheme. Again, we do not find a meaningful connection with the psychological literature.

4.3. Analyses

After viewing a participant’s response to a comparison task, the coder marked the value that the participant identified as larger, marked whether the response was correct or incorrect, and selected from among 11 possible justification codes, which had been generated through constant comparative analysis of pilot data (Corbin & Strauss, 2008). Participants could receive more than one justification code if appropriate to the justification the participant shared. The justification codes fell into one of three broad categories: order-based (4 justification codes), magnitude-based (4 justification codes), or developmental (2 justification codes). If none of the justification codes matched the justification provided, the justification was marked as *other* and the justification was written on the code sheet.

After the data were coded, the responses to each task were organized by student group to identify percentage correct for each task and to identify the number in each group assigned each justification code. The justification codes within each broad way of reasoning were then combined to determine how often the three broad ways of reasoning were used and the degree to which each way of reasoning was used to make a correct comparison. In the following section, we present our findings and trends in the data.

5. Results

In answer to our research questions, we present findings concerning the integer-comparison justifications of college-track 11th graders, 7th graders, 2nd and 4th graders with negatives, and 2nd and 4th graders without negatives. We report broad trends by groups of students in terms of percentages of correct answers on each comparison task. We also report percentages of correct answers by way of reasoning and the prevalence of these ways of reasoning for each student group.

5.1. Performance trends by student group and problem type

Table 1 presents the percentages of correct responses to each comparison item for each group of students. These results indicate clear trends by both student group and problem type. The 2nd and 4th graders without negatives gave few correct answers except when comparing two nonnegative numbers. The 2nd and 4th graders with negatives performed considerably better, with each item answered correctly by the majority of the students. The 7th graders answered almost all comparison tasks correctly, as did the college-track 11th graders.

All 160 students in the sample answered the control item (7 cf. 3) correctly. For each of the other comparisons, marked differences in performance were seen. Comparing a negative to a positive (−7 cf. 3) was the second easiest item, answered correctly by more than one third of 2nd and 4th graders without negatives and by nearly 90% of 2nd and 4th graders with negatives. Comparing a negative number to zero (−9 cf. 0) was considerably more difficult, answered correctly by just more than 20% of 2nd and 4th graders without negatives and just fewer than 80% of 2nd and 4th graders with negatives. The tasks that involved comparing two negatives were most difficult. The comparison of two negative numbers that were close in magnitude (−5 cf. −6) was more difficult than the comparison of two negative numbers that were not close in magnitude (−5 cf. −100). In particular, about 15% of 2nd and 4th graders without negatives and about 60% of 2nd and 4th graders with negatives correctly compared −5 and −100. By contrast, only about 7% of 2nd and 4th graders without negatives and about

Table 1
Percentages of Correct Responses to Comparison Tasks, by Student Group.

Comparison items, ordered by difficulty	Percentage of correct responses			
	2nd and 4th graders without negatives	2nd and 4th graders with negatives	7th graders	11th graders
7 cf. 3	100% (41/41)	100% (39/39)	100% (40/40)	100% (40/40)
−7 cf. 3	36.59% (15/41)	89.74% (35/39)	100% (40/40)	100% (40/40)
−9 cf. 0	21.95% (9/41)	79.49% (31/39)	95% (38/40)	100% (40/40)
−5 cf. −100	14.63% (6/40)	61.54% (24/39)	97.5% (39/40)	97.5% (39/40)
−5 cf. −6	7.32% (3/41)	51.28% (20/39)	97.5% (39/40)	97.5% (39/40)

Table 2
Percentages of Correct Answers by Way of Reasoning (Across All Comparison Problems Involving a Negative Integer).

Student group	Percentages of correct answers by way of reasoning		
	Order	Magnitude	Developmental/other
2nd and 4th graders without negatives	100% (17/17)	36.3% (8/32)	7% (9/128)
2nd and 4th graders with negatives	89.1% (99/111)	20.51% (8/39)	35.2% (6/17)
7th graders	99.3% (142/143)	76.47% (13/17)	100% (8/8)
11th graders	99.12% (113/114)	98.11% (52/53)	88.9% (8/9)[a]

[a] All instances were "other".

Table 3
Numbers and Percentages of Responses by Item That Involved Each Way of Reasoning by 2nd and 4th Graders Without Negatives.

Item	Frequency of way of reasoning by 2nd and 4th without negatives		
	Order	Magnitude	Developmental/other
−7 cf. 3	6 (14.6%)	4 (9.7%)	31 (75.6%)
−9 cf. 0	6 (13.9%)	5 (11.6%)	32 (74.4%)
−5 cf. −100	4 (9.7%)	5 (12.1%)	32 (78%)
−5 cf. −6	1 (2.3%)	8 (19.04%)	33 (78.5%)

50% of 2nd and 4th graders with negatives correctly compared −5 and −6. These results indicate the following hierarchy of problem difficulty for integer-comparison tasks, from easiest to most difficult:

- Positive versus positive
- Positive versus negative
- Negative versus zero
- Negative versus negative with numbers far apart
- Negative versus negative with numbers close together

With a larger set of comparison tasks, this hierarchy could be further refined. For example, had we not asked students to compare both the pair −5 and −6 and the pair −5 and −100 in our interviews, we would not have found the difference in performance on those two subcategories of comparing two negative numbers.

5.2. Prevalence of ways of reasoning and performance by way of reasoning

Table 2 shows the percentages of correct answers obtained by each student group when using each way of reasoning. The numbers of instances of each way of reasoning also show the relative prevalence of each way of reasoning for each student group. We saw noteworthy differences in the ways of reasoning used, depending on both grade level and problem type. Unsurprisingly, most instances (128 of 162) of developmental/other reasoning came from the 2nd and 4th graders without negatives. Order was the most common way of reasoning among 2nd and 4th graders with negatives (occurring in 111 of the responses), and it was used in nearly 90% of the total responses of 7th graders. It was also the most common way of reasoning among the college-track 11th graders, but, at 66.3% of responses, it was considerably less common than among the 7th graders.

Across problems, order-based reasoning clearly led to greater success in integer comparisons for students in all groups (although this difference was negligible for the college-track 11th graders, who answered nearly every problem correctly). The percentages of responses involving order-based reasoning increased dramatically from group to group up to 7th grade but then decreased among the college-track 11th graders. Success rates with order-based reasoning were all high (89% and higher). By contrast, magnitude-based reasoning rarely yielded correct answers for 2nd and 4th graders with or without negatives. Seventh graders used magnitude-based reasoning rarely but were successful in more than 3/4 of their uses. College-track 11th graders used magnitude-based reasoning in about 30% of their responses, and all but one of these responses was correct. Developmental/other reasoning was very common and almost never successful among 2nd and 4th graders without negatives. It was scarcely used by students outside of that group.

Looking at responses to specific items enables us to examine relationships between the particular numbers involved in integer comparisons and the ways of reasoning used by each student group. Table 3 shows the prevalence of each way of reasoning among 2nd and 4th graders without negatives in response to each comparison task. At the level of *way of reasoning*, this group's responses were quite consistent. For each item, about 3/4 of the responses were in the developmental category. Other responses were split more or less equally between order-based reasoning and magnitude-based reasoning, except for the comparison of −5 and −6. For this item, magnitude-based reasoning was considerably more common.

Table 4
Numbers and Percentages of Responses by Item That Involved Each Way of Reasoning by 2nd and 4th Graders With Negatives.

Item	Frequency of way of reasoning by 2nd and 4th without negatives		
	Order	Magnitude	Developmental/other
−7 cf. 3	32 (78%)	5 (12.1%)	4 (9.7%)
−9 cf.0	31 (73.8%)	6 (14.2%)	5 (11.9%)
−5 cf. −100	24 (58.5%)	13 (31.7%)	4 (9.7%)
−5 cf. −6	24 (55.8%)	15 (34.8%)	4 (9.3%)

Table 5
Numbers and Percentages of Responses by Item That Involved Each Way of Reasoning by 7th Graders.

Item	Frequency of way of reasoning by 7th graders		
	Order	Magnitude	Developmental/other
−7 cf. 3	36 (87.8%)	2 (4.8%)	3 (7.3%)
−9 cf. 0	37 (90.2%)	2 (4.8%)	2 (4.8%)
−5 cf. −100	34 (85%)	5 (12.5%)	1 (2.5%)
−5 cf. −6	36 (78.2%)	8 (17.3%)	2 (4.3%)

Table 6
Numbers and Percentages of Responses by Item That Involved Each Way of Reasoning by 11th Graders.

Item	Frequency of way of reasoning by 11th *graders*		
	Order	Magnitude	Developmental/Other
−7 cf. 3	32 (76.2%)	8 (19.0%)	2 (4.8%)
−9 cf. 0	31 (73.8%)	9 (21.4%)	2 (4.8%)
−5 cf. −100	25 (59.1%)	16 (36.4%)	2 (4.6%)
−5 cf. −6	25 (52.1%)	20 (41.7%)	2 (6.3%)

5.3. Prevalence of ways of reasoning by student group

Table 4 shows the prevalence of each way of reasoning among 2nd and 4th graders with negatives in response to each comparison task. For this group, developmental reasoning was rare, at about 10% for each item. We see similar proportions for the comparisons involving one negative integer (−7 vs. 3 and −9 vs. 0) with about three fourths of the students using order-based reasoning. We also see similar proportions for the comparisons of two negative integers (−5 vs. −6 and −5 vs. −100) with more than half of the responses indicating order-based reasoning and about one-third indicating magnitude-based reasoning.

Table 5 shows the prevalence of each way of reasoning among 7th graders in response to each comparison task involving a negative integer. For this group, developmental/other reasoning was very rare, and order-based reasoning was extremely common. As with the 2nd and 4th graders with negatives, magnitude-based reasoning was seen more frequently in comparisons of two negative integers (−5 cf. −6 and −5 cf. −100) than in comparisons involving only one negative integer (−7 cf. 3 and −9 cf. 0).

Table 6 shows the prevalence of each way of reasoning among the college-track 11th graders in response to each comparison task involving a negative integer. As for the 7th graders, developmental/other reasoning was very rare, and order-based reasoning was the most common for all problems. Again, magnitude-based reasoning was seen more frequently in comparisons of two negative integers (−5 cf. −6 and −5 cf. −100) than in comparisons involving only one negative integer (−7 cf. 3 and −9 cf. 0). On each item, magnitude-based reasoning was used considerably more often by the 11th graders than by the 7th graders.

In summary, we found clear differences in students' performance on integer comparisons tasks by student group. Both the college-track 11th graders and the seventh graders appeared to have mastered integer comparisons, 2nd and 4th graders with negatives performed moderately well, and 2nd and 4th graders without negatives understandably tended to answer comparisons involving negative numbers incorrectly. By and large, students using order-based reasoning outperformed those using magnitude-based reasoning. Among the youngest three groups, the prevalence of order-based reasoning increased from group to group when students' familiarity with negative numbers increased; however, it decreased among college-track 11th graders. Among all student groups, magnitude-based reasoning was used more often to compare two negative numbers than to compare a negative with a nonnegative. Magnitude-based reasoning rarely led to correct answers when used by the 2nd and 4th graders with or without negatives, but it was used more effectively by the 7th graders and especially by the college-track 11th graders.

6. Discussion

Our aim in this study was to identify the conceptions students have about comparing integers, ascertain the justifications they offer for comparison claims, and determine which classes of comparison problems are easier or more challenging for students. In our data, we observed clear trends depending on the numbers involved in the comparison task. In the results reported here, students compared integers classified in five categories of pairs: positive versus positive, positive versus negative, negative versus negative with a small difference, negative versus negative with a large difference, and negative versus zero. We analyzed the responses and justifications to the four comparison problems involving negative integers and determined the distribution among categories of ways of reasoning to identify specific trends. We discuss findings in three categories: by looking across problems, looking across grades, and looking across students' ways of reasoning.

6.1. Discussion of trends found in the results

We discuss the observed trends in student reasoning and performance. We then consider implications of the results for educators.

6.1.1. Looking across problems

In results analyzed across problems, we found a clear and consistent pattern of difficulty among the five problem types. According to our results, the order of easiest to most challenging problem type was comparing positive versus positive, positive versus negative, negative versus zero, negative versus negative with a large difference between the two numbers, and negative versus negative with a small difference between the two numbers. This result held for 2nd and 4th graders, regardless of their degree of familiarity with integers. For the 7th and 11th graders, because of ceiling effects, we do not claim to have found differences in difficulty.

We were interested to find that more students correctly compared −5 and −100 than −5 and −6, although both of these comparisons involved two negative integers. We observed that some students used order-based reasoning to compare −5 and −100, even though they used magnitude-based reasoning in comparing −5 and −6. Overall, there were more instances of order-based reasoning for −5 versus −100 and more instances of magnitude-based reasoning for −5 versus −6. These were differences of just a few instances, but they were consistent across all three groups of students. For example, one 2nd grader with negatives said, "Because 6 is greater than 5, −6 should be greater than −5," whereas the same student also claimed, "Negative 5 is a smaller number than negative 100, so it has to be greater. Negative 5 is closer to 1. Negative 100 is away from 1."

In examples like the above, we see evidence of conflict, which may be indicative of a transitional phase between comparing using order and using magnitude. We believe that both order-based and magnitude-based reasoning are generally available to students, given their experiences with the natural numbers. Students with limited experience with negative numbers (as in the 2nd and 4th *with negatives* group) may draw on either or both of these in response to a particular task. Perhaps the difference between −100 and −5 is so great that some students focused on the distance of these numbers from 0 even though that idea was not salient for them when they compared −5 and −6. As noted in the descriptions of the justifications we identified, students can draw either correct or incorrect conclusions when reasoning in terms of either magnitude or order.

Gullick and Wolford (2013) reported a distance effect in the response times of 5th and 7th graders for negative–negative comparisons. This effect was consistent with our results and those of other studies (that pairs of numbers that are close together tend to be more difficult to compare than pairs that are far apart), although our results concern accuracy rather than response time.

6.1.2. Reasoning trends by student group

Data across the three groups of students revealed how students' answers and justifications were distributed among the ways of reasoning. About 3/4 of the 2nd and 4th graders without negatives used developmental/other ways of reasoning across the four comparison questions that included at least one negative integer. Second and 4th graders with or without negatives used order-based reasoning more often to compare a negative integer with a positive integer or with zero (−7 vs. 3 or −9 vs. 0) than to compare two negative integers (−5 vs. −6 or −5 vs. −100). These students were more likely to use magnitude-based reasoning on the latter two problems than on the former two. Students can compare negative numbers with nonnegative numbers by reasoning that negatives are less than zero; however, that reasoning is insufficient for comparing two negatives.

In general, the prevalence of order-based reasoning increased substantially from 2nd and 4th graders without negatives to 2nd and 4th graders with negatives to 7th graders. Interestingly, it was less prevalent in the responses of the college-track 11th graders than for 7th graders.

6.1.3. Performance trends by way of reasoning

Another theme that emerged from our findings is based on students' ways of reasoning and relationships between way of reasoning and correct answers. Among the 2nd, 4th, and 7th graders, order-based reasoning outperformed the other ways of reasoning. For instance, 74% of the 2nd and 4th graders with negatives who used order-based reasoning correctly compared

−5 and −6, whereas only 13% of the students who use magnitude-based reasoning gave the correct answer. Comparable results were found for every comparison problem involving a negative number. Thus, order-based reasoning is associated with correctness on these tasks.

We caution readers to avoid the conclusion that order-based reasoning is the only valuable way of reasoning about integers. Integer comparisons are a particular type of task, and order-based reasoning works well for this type of task because the conventional meanings of *greater than* and *less than* are based on order, as opposed to magnitude. However, as noted earlier, researchers have previously identified productive ways in which students can and do make use of magnitude-based reasoning in integer arithmetic. Furthermore, we found that the college-track 11th graders used magnitude-based reasoning more frequently than did the 7th graders, and they almost always used it correctly. The college-track 11th graders as a group exhibited flexible use of order-based and magnitude-based reasoning, and their performance on the set of tasks was masterful. Thus, our results indicate that ultimately order-based reasoning does not trump magnitude-based reasoning. Rather, students who have developed relatively sophisticated understandings of integers are able to effectively use both ways of reasoning.

6.2. Instructional implications

According to Steffe (1991), "As mathematics educators, we have a choice between using mathematics of children or conventional school mathematics as the basis on which to teach mathematics. Choosing the former is a fundamental requirement of constructivism for mathematics education" (p. 181). To base mathematics teaching on the mathematics of children, we must first discover and make sense of children's mathematics. We interviewed 160 students in Grades 2, 4, 7, and 11 and analyzed their responses to identify students' justifications and underlying ways of reasoning about integer comparisons. The findings illuminate the field's understanding of children's mathematics in this domain. Instructional implications follow.

We have seen that for younger students, order-based reasoning tends to be used more successfully for comparing integers because it aligns with the mathematical community's conventional meanings for inequalities involving negative numbers. In terms of implications for teaching, we see the importance of recognizing that both order and magnitude are familiar ideas for students in the context of whole numbers. When students come to integer instruction, they have had extensive experience reasoning about whole numbers in terms of both magnitude and order. They can see 6 as being greater than 5 in the sense that 6 things is more than 5 things. They can also see 6 as being greater than 5 in the sense that 6 comes after 5 in the counting sequence and is to the right of 5 on the number line. In comparing negative numbers, the question is which of these ways of reasoning will students invoke?

Our recommendation is to create the opportunity for contrasting ways of reasoning about comparing negatives to arise and be discussed. We do not believe that one of these ways makes sense and the other does not. On the contrary, both are sensible, but they conflict. So, in the interest of communication, conventional meanings for notations and terms related to comparisons must be established. The mathematical community uses the symbols < and > and the terms *less than* and *greater than* in ways that are consistent with order-based reasoning. That is, the expression $-5 > -6$ can be taken to mean that −5 is to the right of −6 on the number line. In that sense, −5 is greater. At the same time, we can appreciate the sense in which −5 is less than −6. There is also a conventional notation that we can use to express this idea. By writing $|-5| < |-6|$, we specify that we are focusing on the magnitudes of the numbers.

For those students who are reasoning about integer comparisons in terms of magnitude, a teacher's reminder that numbers further to the right on the number line are greater may be of no help. Instead, we suggest that teachers acknowledge that such students are making sense. Their reasoning is not flawed; rather, two reasonable possibilities exist, and a conventional interpretation is needed in the interest of precise communication (NGA & CCSSO, 2010). Rather than tell students to compare negative numbers in one way and not the other, teachers can help them to become fluent in the written and oral language of mathematics and to learn to express their ideas in ways that others will understand.

The results from the college-track 11th graders underscore the point that order-based reasoning is not simply superior to magnitude-based reasoning when it comes to integer comparisons. There are ways of using magnitude-based reasoning to compare integers that are both sensible and correct. The college-track 11th graders selectively used order-based or magnitude-based reasoning depending on the numbers involved in the comparison task, and their answers were overwhelmingly correct. For example, an 11th grader using magnitude-based reasoning to compare −5 with −6 stated, "I know that the littler number in negative, like I said, it's opposite. Because 5 would have been less than 6, but if it is negative, then 5 is bigger than 6." Other 11th graders invoked contexts such as debt to reason sensibly and correctly compare negative integers. They also flexibly invoked order-based comparisons, such as referencing relative positions on a number line. This result indicates that a mature understanding of integer comparisons integrates both order and magnitude. What students may initially experience as conflicting ways of reasoning about integers can eventually find harmony, and students can learn to reason correctly and flexibly about integer comparisons.

6.3. Reflections on the literature

We relate our research to the two literature bases that informed this study. We focus especially on connections to the mathematics education literature and contrast with the psychological literature.

6.3.1. Connections to mathematics education literature

In the relevant mathematics education literature order-based and magnitude-based reasoning have been emphasized and particular strategies and conceptions identified that students may use when reasoning about integer meanings, comparisons, or operations. Through our coding scheme, we organized students' justifications for integer comparisons into a coherent framework. Whereas Bofferding (2014), in her scheme of mental models, emphasizes fine-grained developmental levels in students' early conceptions of integers, our framework consists of a variety of possible justifications organized according to broader ways of reasoning. We found that the developmental/other category was typical of students who had little or no familiarity with negative numbers. Beyond this distinction, however, our categories are non-hierarchical. In fact, many students competently used multiple justifications within both categories of order-based and magnitude-based reasoning.

6.3.2. Contrasts with the psychological literature

Some experimental psychologists have taken an interest in the underlying mental representations or processes that people may use when comparing integers. Our study contrasts with this research tradition in several ways. The psychological research has focused on competent performance, especially the performance of college students, in response to integer-comparison tasks. Some researchers have even provided real-time feedback on the correctness of participants' responses, thus correcting them. By contrast, we investigated the performance and reasoning of K-12 students, and we were interested in trends in their performance and reasoning, whether or not their responses were correct. Whereas the psychological studies have focused on response time, we considered response time irrelevant to our investigation.

In testing the phylogenetic, ontogenetic, and other hypotheses, researchers in the psychological tradition seek a single explanation for how people compare integers. By contrast, we find that students—even those who perform competently—report reasoning in a variety of ways about integer comparisons. Thus, in our view, reasoning about integer comparisons has no single explanation. Different students reason in different ways about the tasks, and the same students may reason in a variety of ways. Teachers need to be aware of these different ways of reasoning and particular justifications to support students in making sense of integers. Thus, frameworks such as ours and that of Bofferding (2014), like other frameworks of students' mathematical thinking (e.g., Carpenter et al., 1999), have the potential to be shared with teachers and applied productively in practice.

7. Conclusions

Our research was informed by previous studies that helped to sensitize us to the distinction between magnitude- and order-based reasoning and raised questions for us regarding students' reasoning about integer comparisons. We have contributed a systematic study of the justifications and ways of reasoning for integer comparisons of four groups of students—2nd and 4th graders without negatives, 2nd and 4th graders with negatives, 7th graders, and 11th graders. These groups provide a cross-sectional look at students' reasoning about integers at four relevant points in mathematical education: (a) at a time in the elementary school years when they know very little or nothing about negative numbers, (b) at a time in the elementary school years when they have some meaning for negative numbers but have not experienced integer instruction, (c) in middle school, after substantial integer instruction, and (d) in high school, for those on a successful mathematics track. Trends in students' reasoning across these three groups reveal differences related to the development of students' knowledge of integers.

We have contributed a framework consisting of three broad categories of reasoning about integer comparisons, complete with the specific justifications belonging to each category. This framework was not determined a priori; instead, it is grounded in the responses of the 160 students who participated in this study. The framework was further informed by additional interviews, instructional experiences, discussions with colleagues, and previous research in this area. Such knowledge frameworks are essential to a mathematics education research literature that has the power to profoundly inform instruction.

Another of Steffe's (1991) requirements of constructivism for mathematics education is to choose "interpreting children's mathematical activity in learning environments through interactive communication, *making qualitative distinctions*" over "focusing on the result or product of their activity" (p. 185, emphasis added). Integer-comparison tasks have only three possible products, or answers—less than, greater than, or equal to—and students' achievement may be assessed on the basis of frequency of correct answers. Although these answers provide some information, students' reasoning is revealed in the justifications they offer. The framework we have proposed provides qualitative distinctions to inform interpretations of students' ways of reasoning. Frameworks of this sort have proven to be valuable to teachers and to have the power to transform mathematics teaching and learning (e.g., Carpenter et al., 1999). We offer this framework as a tool for mathematics teachers and teacher educators to inform interpretations of students' reasoning about integer comparisons and integers more broadly (Bishop et al., 2014) and to shape instructional goals to focus on ways of reasoning.

Acknowledgments

We thank the cooperating teachers and students without whom this study would have been impossible. We also thank the editor and anonymous reviewers for their thoughtful feedback on previous drafts of this manuscript.

References

Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The Elementary School Journal*, *93*(4), 373–397.

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

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

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

Bruno, A., & Martinón, A. (1999). The teaching of numerical extensions: The case of negative numbers. *International Journal of Mathematical Education in Science and Technology*, *30*, 789–809.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (1999). *Children's mathematics: Cognitively guided instruction*. Portsmouth, NH: Heinemann.

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

Corbin, A., & Strauss, J. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: Sage.

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

Fischer, M. H. (2003). Cognitive representation of negative numbers. *Psychological Science*, *14*(3), 278–282.

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

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*, 171–192.

Ganor-Stern, D., & Tzelgov, J. (2008). Negative numbers are generated in the mind. *Experimental Psychology*, *55*(3), 157–163.

Gullick, M. M., & Wolford, G. (2013). Understanding less than nothing: Children's neural response to negative numbers shifts across age and accuracy. *Frontiers in Psychology*, *4*(584), 1–17.

Hativa, N., & Cohen, D. (1995). Self-learning of negative number concepts by lower division elementary students through solving computer-provided numerical problems. *Educational Studies in Mathematics*, *28*, 401–431.

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

Huber, S., Cornelsen, S., Korbinian, M., & Nuerk, H.-C. (2015). Toward a model framework of generalized parallel componential processing of multi-symbol numbers. *Journal of Experimental Psychology*, *41*(3), 732–745.

Krajcsi, A., & Igács, J. (2010). Processing negative numbers by transforming negatives to positive range and by sign shortcut. *European Journal of Cognitive Psychology*, *22*(7), 1021–1038.

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

Moses, R. P., & Cobb, C. E. (2002). *Radical equations: Civil rights from Mississippi to the Algebra Project*. Boston, MA: Beacon Press.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Author. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.

Parnes, M., Berger, A., & Tzelgov, J. (2012). Brain representations of negative numbers. *Canadian Journal of Experimental Psychology*, *66*(4), 251–258.

Peled, I., & Carraher, D. W. (2007). Signed numbers and algebraic thinking. In J. Kaput, D. Carraher, & M. Blanton (Eds.), *Algebra in the early grades* (pp. 303–327). Mahwah, NJ: Erlbaum (now Taylor & Francis).

Peled, I. (1991). Levels of knowledge about signed numbers: Effects of age and ability. In F. Furinghetti (Ed.), *Proceedings of the 15th annual conference of the International Group for the Psychology of Mathematics Education* (vol. 3) (pp. 145–152). Assisi, Italy: Conference Committee.

Steffe, L. P., & Olive, J. (2010). *Children's fractional knowledge*. New York, NY: Springer.

Steffe, L. P. (1991). The constructivist teaching experiment: Illustrations and implications. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 177–194). New York: Kluwer.

Steffe, L. P. (2002). A new hypothesis concerning children's fractional knowledge. *Journal of Mathematical Behavior*, *20*(3), 267–307. http://dx.doi.org/10.1016/s0732-3123(02)00075-5

Steffe, L. P. (2004). On the construction of learning trajectories of children: The case of commensurate fractions. *Mathematical Thinking and Learning*, *6*(2), 129–162. http://dx.doi.org/10.1207/s15327833mtl0602_4

Stephan, M., & Akyuz, D. (2012). A proposed instructional theory for integer addition and subtraction. *Journal for Research in Mathematics Education*, *43*(4), 428–464. http://dx.doi.org/10.5951/jresemathe- duc.43.4.28

Varma, S., & Schwartz, D. L. (2011). The mental representation of integers: An abstract-to-concrete shift in the understanding of mathematical concepts. *Cognition*, *121*, 363–385.

Vlassis, J. (2004). Making sense of the minus sign or becoming flexible in negativity. *Learning and Instruction*, *14*(5), 469–484. http://dx.doi.org/10.1016/j.learninstruc.2004.06.012

Whitacre, I. M., Bishop, J. P., Lamb, L. L. C., Philipp, R. A., Schappelle, B. P., & Lewis, M. (2011). Integers: History, textbook approaches, and children's productive mathematical intuitions. In L. R. Wiest, & T. Lamberg (Eds.), *Proceedings of the 33rd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 913–920). Reno, NV: University of Reno.

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

Whitacre, I., Bouhjar, K., Bishop, J. P., Philipp, R. A., Schappelle, B. P., & Lamb, L. L. (2016). Regular numbers and mathematical worlds. *For the Learning of Mathematics*, *36*(2), 21–26.

Wilcox, V. (2008). Questioning zero and negative numbers. *Teaching Children Mathematics*, *15*(4), 202–206.