

MORE AND LESS: LANGUAGE SUPPORTS FOR LEARNING NEGATIVE NUMBERS

Laura Bofferding
Purdue University
lbofferd@purdue.edu

Sherri Farmer
Purdue University
farmer10@purdue.edu

The language that students use with whole numbers can be insufficient when learning integers. This is often the case when children interpret addition as “getting more” or “going higher.” In this study, we explore whether instruction on mapping directed magnitudes to operations helps 88 second graders and 70 fourth graders solve addition and subtraction problems with negative numbers. Further we explore to what extent having prior training with directed magnitude language (as opposed to just more and less language, without a direction specified) prepares students to benefit from the instruction. Our data shows that students, regardless of which language training they had, improved, and second graders, even with less initial knowledge, were able to make the same gains as fourth graders, suggesting that having initial exposure to negatives earlier could help students reach proficiency by the time the standards expect it.

Keywords: Number Concepts and Operations, Elementary School Education, Instructional Activities and Practices

Because subjects are usually taught separately in elementary school, there is often a boundary between language and mathematics instruction. Yet the National Council for Teachers of Mathematics (NCTM, 2000) standards call for students to “use the language of mathematics to express mathematical ideas precisely” (p. 60). Precise language can clarify ideas; whereas, vague language can create confusion. Unfortunately, the language that students use with whole numbers can be insufficient when they learn new numbers, such as integers. This is often the case when children interpret addition as “getting more” or “going higher.” Consider, for example, the following second grade student (B05) explaining why $-1 + 8 = -9$: “Because negative one, and then eight *more* is negative nine.” The student’s statement includes language that highlights the operation (more = addition) and magnitude (8). However, “more” in integer operations has multiple meanings. The problem involves getting *more positive* (counting up the number sequence in the positive direction), an operation with a directed magnitude (adding positive 8), as opposed to getting *more negative*. Therefore, finding ways to help students interpret and apply directed magnitude language to comparisons and operations with negative integers is necessary as they transition from work with whole numbers to integers.

Theoretical Framework

Before the introduction of negative numbers, students learn that “more” corresponds to an increase in magnitude, which corresponds to counting up the number sequence (moving to the right on the number line), and maps onto addition (Case, 1996). With the inclusion of negative numbers, an additional level of specificity is needed. Students need to learn that an increase *in the positive direction* is getting “more positive”, which corresponds to moving right on the number line and maps onto adding a positive number. Similarly, *an increase in the negative direction* is getting “more negative”, which corresponds to moving left on the number line and maps onto adding a negative number. Similar relations exist for subtraction, except instead of an increase in a direction, there is a decrease in a direction (getting less positive or less negative).

Students’ understanding of comparisons with more and less has been explored both from a language perspective and a mathematical perspective. Language studies indicate that, initially, comparisons with positive or unmarked adjectives are easier for students to answer than negative or marked ones. Therefore, determining which object is *higher* is easier than *lower* (Smith, Rattermann,

& Sera, 1988), and determining which set of objects or glass of liquid has *more* is easier than *less* (Palermo, 1973). However, there is still some debate as to whether this happens because positive adjectives are used more or because they are conceptually easier to understand (e.g., Ryalls, 2000).

Studies from a mathematical perspective focus on exploring students' responses to questions about which of two numbers is more. Students tend to categorize numbers as large and small. For positive number comparisons, when two numbers belong to different categories defined by the students (e.g., large and small), young children do better on the comparisons than when both numbers are part of the same category (e.g., two numbers they consider large) (Murray & Mayer, 1988). When comparing two negatives, experienced sixth graders are faster when the two numbers are farther apart than when they are closer. However, there are no differences when they compare a positive and negative number, suggesting they use a rule that positives are more than negatives regardless of how far apart the numbers are (Varma & Schwartz, 2011).

Aside from teaching students the rule that positives are greater than negatives, using questions that include a context could make the desired value or point of comparison more explicit. For example, if integer problems were presented in a golf context, negative numbers would be better than positive numbers; whereas, this would be reversed for many video games or board games, where the goal is to achieve a higher positive score. The desired value is also clearer with temperature (e.g., which temperature is hottest vs. coldest)? To get the same level of obviousness with numbers there would need to be a move away from the magnitude questions (where positive is the assumed reference category) to directed magnitude questions (where we explicitly say which type of value - positive or negative - is the reference category).

Knowing which direction "more" relates to (as opposed to assuming more always means more positive) is especially important when adding and subtracting negative integers. Students who have learned about negatives know that positive numbers are considered more than negative numbers and that the question, "Which is more?" is asked from a positive perspective. Therefore, they might be more receptive to a focus on the language connecting integer addition and subtraction to directed magnitudes. This would also align with the Common Core Standards for Mathematics' recommendation that integers be introduced in the sixth grade (National Governor's Association Center for Best Practices & Council of Chief State School Officers, 2010). On the other hand, students with little experience with negatives often determine which number is more based on absolute value (Bofferding, 2014). Given young students' willingness to consider "large negatives" as more, they might be more receptive to such a focus on language. Further, if current magnitude language, emphasized in whole number instruction, is limiting, it is possible that using directed magnitude language in early elementary would help students build more cohesive conceptions of "more" and "less".

Research Questions

Based on the issues described above, we explore whether instruction on mapping directed magnitudes to operations helps second and fourth graders solve addition and subtraction problems with negative numbers. Further we explore to what extent having prior training with directed magnitude language (as opposed to just *more* and *less* without a direction specified) prepares students for the instruction. Specifically, we investigate the following research question and sub-questions:

How does instruction mapping directed magnitude language to operations with integers affect students' learning of integer comparisons, addition, and subtraction?

- (a) To what extent does having a background understanding of language related to magnitude (e.g., more, lower) versus directed magnitude (e.g., more positive, less negative) influence learning?
- (b) How do gains differ for students in second grade versus fourth grade?

Methods

Participants and Setting

This study took place at two public schools in the same rural Mid-western district. Students were recruited from all of the second and fourth grades, resulting in 88 second-grade, and 70 fourth-grade participants. The district reports 32.2% English-language learners and 75.2% as qualifying for free or reduced-lunch.

Design and Materials

The study involved a pretest, random assignment to language group (magnitude or directed magnitude), three small group sessions, individual language training, whole class instruction, and a posttest. We describe each of these below.

Pretest. Students individually completed a written pretest during a whole class session. Students not in the study had their tests returned to their teacher or shredded. The pretest consisted of filling in the missing numbers on a number path, value comparisons with different language (n=24), feeling-related questions (n=4), and integer addition and subtraction problems (n=13) (see Table 1 for examples).

Table 1: Example Items from the Pretest

Question Type	Examples																																																		
Integer Order	<p>Fill in the missing numbers before 4 and after 6.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px; text-align: center;">4</td> <td style="width: 20px; height: 20px; text-align: center;">5</td> <td style="width: 20px; height: 20px; text-align: center;">6</td> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> </tr> </table>								4	5	6																																								
							4	5	6																																										
Value Comparisons	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; vertical-align: top;"> <p>Circle the temperature that is... ...hottest...most hot...least cold.</p> <table border="1" style="width: 100%; text-align: center;"> <tr> <td>2 °F</td> <td>3 °F</td> <td>8 °F</td> <td>none</td> </tr> <tr> <td>0 °F</td> <td>-6 °F</td> <td>-1 °F</td> <td>none</td> </tr> <tr> <td>-4 °F</td> <td>-2 °F</td> <td>3 °F</td> <td>none</td> </tr> <tr> <td>5 °F</td> <td>-10 °F</td> <td>-8 °F</td> <td>none</td> </tr> <tr> <td>-6 °F</td> <td>-2 °F</td> <td>-3 °F</td> <td>none</td> </tr> <tr> <td>-5 °F</td> <td>-4 °F</td> <td>-7 °F</td> <td>none</td> </tr> </table> </td> <td style="width: 50%; vertical-align: top;"> <p>Circle the temperature that is... ...coldest...most cold...least hot.</p> <table border="1" style="width: 100%; text-align: center;"> <tr> <td>2 °F</td> <td>6 °F</td> <td>4 °F</td> <td>none</td> </tr> <tr> <td>-1 °F</td> <td>0 °F</td> <td>-6 °F</td> <td>none</td> </tr> <tr> <td>3 °F</td> <td>-4 °F</td> <td>-2 °F</td> <td>none</td> </tr> <tr> <td>-8 °F</td> <td>5 °F</td> <td>-10 °F</td> <td>none</td> </tr> <tr> <td>-3 °F</td> <td>-6 °F</td> <td>-2 °F</td> <td>none</td> </tr> <tr> <td>-7 °F</td> <td>-5 °F</td> <td>-4 °F</td> <td>none</td> </tr> </table> </td> </tr> </table>	<p>Circle the temperature that is... ...hottest...most hot...least cold.</p> <table border="1" style="width: 100%; text-align: center;"> <tr> <td>2 °F</td> <td>3 °F</td> <td>8 °F</td> <td>none</td> </tr> <tr> <td>0 °F</td> <td>-6 °F</td> <td>-1 °F</td> <td>none</td> </tr> <tr> <td>-4 °F</td> <td>-2 °F</td> <td>3 °F</td> <td>none</td> </tr> <tr> <td>5 °F</td> <td>-10 °F</td> <td>-8 °F</td> <td>none</td> </tr> <tr> <td>-6 °F</td> <td>-2 °F</td> <td>-3 °F</td> <td>none</td> </tr> <tr> <td>-5 °F</td> <td>-4 °F</td> <td>-7 °F</td> <td>none</td> </tr> </table>	2 °F	3 °F	8 °F	none	0 °F	-6 °F	-1 °F	none	-4 °F	-2 °F	3 °F	none	5 °F	-10 °F	-8 °F	none	-6 °F	-2 °F	-3 °F	none	-5 °F	-4 °F	-7 °F	none	<p>Circle the temperature that is... ...coldest...most cold...least hot.</p> <table border="1" style="width: 100%; text-align: center;"> <tr> <td>2 °F</td> <td>6 °F</td> <td>4 °F</td> <td>none</td> </tr> <tr> <td>-1 °F</td> <td>0 °F</td> <td>-6 °F</td> <td>none</td> </tr> <tr> <td>3 °F</td> <td>-4 °F</td> <td>-2 °F</td> <td>none</td> </tr> <tr> <td>-8 °F</td> <td>5 °F</td> <td>-10 °F</td> <td>none</td> </tr> <tr> <td>-3 °F</td> <td>-6 °F</td> <td>-2 °F</td> <td>none</td> </tr> <tr> <td>-7 °F</td> <td>-5 °F</td> <td>-4 °F</td> <td>none</td> </tr> </table>	2 °F	6 °F	4 °F	none	-1 °F	0 °F	-6 °F	none	3 °F	-4 °F	-2 °F	none	-8 °F	5 °F	-10 °F	none	-3 °F	-6 °F	-2 °F	none	-7 °F	-5 °F	-4 °F	none
<p>Circle the temperature that is... ...hottest...most hot...least cold.</p> <table border="1" style="width: 100%; text-align: center;"> <tr> <td>2 °F</td> <td>3 °F</td> <td>8 °F</td> <td>none</td> </tr> <tr> <td>0 °F</td> <td>-6 °F</td> <td>-1 °F</td> <td>none</td> </tr> <tr> <td>-4 °F</td> <td>-2 °F</td> <td>3 °F</td> <td>none</td> </tr> <tr> <td>5 °F</td> <td>-10 °F</td> <td>-8 °F</td> <td>none</td> </tr> <tr> <td>-6 °F</td> <td>-2 °F</td> <td>-3 °F</td> <td>none</td> </tr> <tr> <td>-5 °F</td> <td>-4 °F</td> <td>-7 °F</td> <td>none</td> </tr> </table>	2 °F	3 °F	8 °F	none	0 °F	-6 °F	-1 °F	none	-4 °F	-2 °F	3 °F	none	5 °F	-10 °F	-8 °F	none	-6 °F	-2 °F	-3 °F	none	-5 °F	-4 °F	-7 °F	none	<p>Circle the temperature that is... ...coldest...most cold...least hot.</p> <table border="1" style="width: 100%; text-align: center;"> <tr> <td>2 °F</td> <td>6 °F</td> <td>4 °F</td> <td>none</td> </tr> <tr> <td>-1 °F</td> <td>0 °F</td> <td>-6 °F</td> <td>none</td> </tr> <tr> <td>3 °F</td> <td>-4 °F</td> <td>-2 °F</td> <td>none</td> </tr> <tr> <td>-8 °F</td> <td>5 °F</td> <td>-10 °F</td> <td>none</td> </tr> <tr> <td>-3 °F</td> <td>-6 °F</td> <td>-2 °F</td> <td>none</td> </tr> <tr> <td>-7 °F</td> <td>-5 °F</td> <td>-4 °F</td> <td>none</td> </tr> </table>	2 °F	6 °F	4 °F	none	-1 °F	0 °F	-6 °F	none	3 °F	-4 °F	-2 °F	none	-8 °F	5 °F	-10 °F	none	-3 °F	-6 °F	-2 °F	none	-7 °F	-5 °F	-4 °F	none		
2 °F	3 °F	8 °F	none																																																
0 °F	-6 °F	-1 °F	none																																																
-4 °F	-2 °F	3 °F	none																																																
5 °F	-10 °F	-8 °F	none																																																
-6 °F	-2 °F	-3 °F	none																																																
-5 °F	-4 °F	-7 °F	none																																																
2 °F	6 °F	4 °F	none																																																
-1 °F	0 °F	-6 °F	none																																																
3 °F	-4 °F	-2 °F	none																																																
-8 °F	5 °F	-10 °F	none																																																
-3 °F	-6 °F	-2 °F	none																																																
-7 °F	-5 °F	-4 °F	none																																																
Feelings Comparisons	<table style="width: 100%;"> <tr> <td style="width: 50%; vertical-align: top;"> <p>Circle the face that is least sad.</p> <div style="display: flex; justify-content: space-around;">    </div> </td> <td style="width: 50%; vertical-align: top;"> <p>Draw a face that is less happy than this face.</p> <div style="text-align: center;">  </div> </td> </tr> </table>	<p>Circle the face that is least sad.</p> <div style="display: flex; justify-content: space-around;">    </div>	<p>Draw a face that is less happy than this face.</p> <div style="text-align: center;">  </div>																																																
<p>Circle the face that is least sad.</p> <div style="display: flex; justify-content: space-around;">    </div>	<p>Draw a face that is less happy than this face.</p> <div style="text-align: center;">  </div>																																																		
Addition & Subtraction	<table style="width: 100%;"> <tr> <td style="width: 50%; text-align: center;"> $4 + -5; -1 + 8; -6 + -4$ </td> <td style="width: 50%; text-align: center;"> $1 - 4; -6 - -9; 5 - -3$ </td> </tr> </table>	$4 + -5; -1 + 8; -6 + -4$	$1 - 4; -6 - -9; 5 - -3$																																																
$4 + -5; -1 + 8; -6 + -4$	$1 - 4; -6 - -9; 5 - -3$																																																		

Students were stratified within each grade level based on their performance on the pretest, and then students from each strata were randomly assigned to one of two language groups: a control group that used traditional magnitude language or an experimental group that used directed magnitude language.

Small group sessions. With 2-3 students from their group, students participated in three, 15-minute sessions. The initial session consisted of three activities. The first activity was a game where

students were asked to find matching integers. Cards were arranged in an array, and students turned them over one at a time, seeking the matching value. Researchers asked questions, such as, “How are they (the numbers) the same?” or “How are they different?” in an effort to focus attention on the positive or negative values. During the second activity, students were told that negative numbers are used below zero. Students then explored integer order by filling in spaces on a number path marked only with zero. Using their completed number path as a reference, students played a card game for their third activity. Their goal was to collect and trade integer cards in an effort to get a card hand with three integers in a row (e.g., -2, -1, 0).

The second and third small group sessions both involved playing a game where students moved on a number path labeled -10 to 10. The language of the magnitude group remained consistent with the use of operating with magnitudes. Therefore, they moved “more or less” or “higher or lower” a certain amount; whereas, the directed magnitude group used language consistent with operating with directed magnitudes. They moved “more positive or less positive” and “less negative or more negative” a certain amount.

Training. During their training session, students completed four phrasing-type rounds, with up to twenty-eight comparisons per round. The control group received training on magnitude comparisons (phrasing: Which is more? Which is less? Which is higher? Which is lower?), and the experimental group received training on directed magnitude comparisons (phrasing: Which is more positive? Which is less positive? Which is more negative? Which is less negative?). Students continued only until they got seven consecutively correct for each phrasing type. They were then asked to order a set of non-consecutive integers according to the above-mentioned phrasing (e.g., from most positive to most negative). Both groups then sorted a set of integers into categories three times: high/low, less/more, and positive/negative piles.

Whole-class instruction. In each of the classrooms, two researchers presented a 30-minute lesson to all students using the directed magnitude language. To introduce the idea of moving on a continuum with directed magnitudes, the researchers showed an emotion-continuum of faces from very sad to very happy with a neutral face in the middle. They then read a story about a boy who got more happy, more sad, less happy, and less sad. Students helped show how the boy’s mood changed on the emotion-continuum at the front of the room. After this introduction, students were told that addition means moving *more in a particular direction* (either more positive or more negative) and that subtraction means moving *less in a particular direction* (either less positive or less negative). The researchers then did a series of problems with the class to reinforce the mapping of the directed magnitude language onto the operations. Problems were designed as number strings (DiBrienza & Shevell, 1998; Lampert, Beasley, Ghouseini, Kazemi, & Franke, 2010), and students participated using their own number path and sticker to move as they solved the problems. Table 2 lists the typical order of the problems used.

Table 2: Problems Included in the Class Number Strings

1) $4 + 3$	4) $2 - 6$	7) $-1 - -7$
2) $3 + -4$	5) $2 - -6$	Extra, if time:
3) $-3 + 4$	6) $-6 - 2$	$-7 - -1; -5 + -3; -5 + 3$

Posttest. The posttest was largely the same as the pretest, except for a few additions and changes. We only re-asked the most difficult temperature comparison questions from the pretest ($n=12$). We also changed the feeling comparisons to allow students to pick 1 of 5 faces (very happy, happy, neutral, sad, very sad). Finally, we added a small selection of transfer questions, such as solving for a missing addend, subtrahend, or minuend.

Analysis

We began analysis with a multivariate ANOVA on the difference in temperature comparison scores and the difference in arithmetic scores for grade and condition (including only items present on both pretest and posttest). Then, we conducted a 2 (second grade vs. fourth grade) x 2 (control vs. experimental) x 2 (pretest vs. posttest) x 2 (temperature comparisons vs. arithmetic) repeated measures ANOVA to learn more about potential differences. To better understand students' difficulties with the arithmetic problems, we further analyzed the results by looking at common responses students gave. For each of the arithmetic problems on the pretest and posttest, we calculated the percentage of students who provided a specific numerical response. For example, we looked at the percentage of second and fourth graders who, for $-4 + -6$, answered 10, -10, 2, -2, or 0. Further analysis of the temperature comparisons can be found elsewhere (Bofferding & Farmer, 2016).

Results

Gains from Pretest to Posttest

The results of the multivariate ANOVA were not significant for condition or grade. Therefore, any gains made were statistically similar for students regardless of whether they had the training with magnitude or directed magnitude language, and regardless of whether they were in second grade or fourth grade. Based on the repeated measures ANOVA, there was a significant interaction between test and item type, $F(1, 154) = 9.11, p = .003$. Students did significantly better on the temperature comparisons and arithmetic problems on the posttest than they did on the pretest. Further, on average, performance was significantly higher on the temperature comparisons than the arithmetic problems, $F(1, 154) = 83.27, p < .001$. Students had an average gain of 2.15 on the temperature comparisons versus 1.15 on the arithmetic problems. Fourth graders did start out performing significantly higher than second graders on both items, based on an item type by grade interaction, $F(1, 154) = 17.55, p < .001$; however, as mentioned previously both fourth and second graders made similar gains (see Table 3).

Table 3: Pretest, Posttest, and Gain Scores for Second and Fourth Graders

Grade	Temperature Comparisons Average Correct (SD)			Integer Arithmetic Problems Average Correct (SD)		
	Pretest n=12	Posttest n=12	Gain	Pretest n=13	Posttest n=13	Gain
2 nd (n=88)	2.86 (2.97)	5.38 (3.66)	2.52	2.52 (1.62)	3.61 (2.41)	1.09
4 th (n=70)	6.59 (3.54)	8.37 (3.77)	1.78	4.04 (2.06)	5.24 (2.74)	1.20

Although students as a group made significant gains, we investigated the gains further by running analyses for each grade separately. For fourth graders, there was no significant interaction between test and item type, so their gains from pretest to posttest are only significant when averaging over both item types, $F(1, 69) = 39.83, p < .001$. Averaging over both tests, they also did significantly better on the temperature comparisons than the arithmetic problems, $F(1, 69) = 66.58, p < .001$. For second graders, there was a significant interaction between test and item type, $F(1, 87) = 12.30, p = .001$, with performance on both the temperature comparisons and arithmetic problems increasing from the pretest to the posttest. These results suggest that in the overall analysis, the test by item interaction was largely driven by the second graders.

Arithmetic Problems Involving Negatives

Students' performance on the arithmetic problems involving negatives varied widely, and their gains on these items were not as large partly because many students got a problem correct on the pretest without knowing about negative numbers, and then got the same problem incorrect on the posttest because they were paying attention to the negative and trying to use it. For example, when solving $-2 - -6$ on the pretest, 31% of second graders correctly answered "4" as opposed to 16% of fourth graders. Several second graders who answered "4" also solved $-4 - -3 = 1$, as if all numbers were positive and $4 - -5 = 1$, as if the problem was $5 - 4 = 1$. Therefore, they likely got $-2 - -6$ correct because they solved the problem as $6 - 2 = 4$. In contrast, on the posttest, B01 (who answered 0 on the pretest), explained why the answer was 4 as follows: "So there's a take away, and it's negative. You had to go farther from the negative because it's in the positive." This student had training with the directed magnitude language and used it to reason about the answer, correctly answering all but one of the arithmetic problems.

Both grade levels had 10%-16% of students gain on solving $1 - 4$ and $6 - 8$, suggesting they got better at counting into the negatives. However, 29% - 44% of students at each grade level still solved the problems as $4 - 1$ or $8 - 6$ on the posttest. Both grade levels showed a stronger focus on negatives on the posttest than on the pretest, even if they did not get correct answers. Second graders made a large gain on answering $-6 + -4 = -10$. On the pretest, 16% of them correctly answered this problem, while 34% of them answered "10", as if the numbers were positive. On the posttest, 38% of them correctly answered the problem, and their most prevalent incorrect response changed to "-2". Fourth graders made the largest gain on solving $5 + -2 = 3$ (from 27% correct to 49% correct). On the pretest, 29% of them incorrectly answered "7", but on the posttest, their most common incorrect response was "-7".

Although there were several instances where the fourth graders made larger gains on certain problem types than the second graders, the second graders consistently performed better and made larger gains than fourth graders on problems where they had to subtract a negative number from a positive number (see Table 4). Especially for $5 - -3$, students in both grades solved the problem as $5 - 3 = 2$ on the pretest. Although the same percent of fourth graders still answered this way on the posttest, a lower percent of second graders did so.

Table 4: Percent of Students Providing Different Answers to Integer Questions

4 - -5	Second		Fourth		Change in %	
	Pretest	Posttest	Pretest	Posttest	2 nd Grade	4 th Grade
Correct: 9	2%	15%	0%	6%	+13%	+6%
-9	6%	5%	10%	20%	-1%	+10%
1	25%	14%	10%	16%	-15%	+6%
-1	19%	32%	53%	46%	+13%	-7%
0	25%	17%	10%	3%	-8%	-7%
5 - -3	Second		Fourth		Change in %	
	Pretest	Posttest	Pretest	Posttest	2 nd Grade	4 th Grade
Correct: 8	3%	15%	1%	0%	+12%	+9%
2	59%	47%	49%	49%	-12%	0%
-2	6%	13%	19%	14%	+7%	-5%
0	1%	2%	1%	1%	+1%	0%

Conclusions & Implications

Overall, students in the study benefitted from instruction on mapping the language of directed magnitude onto the operations, regardless of which language they had training on. This suggests that the level of explicitness in the lesson was beneficial. Further, the differences between groups may have been reduced based on this common instruction. Therefore, future work should investigate students' knowledge just prior to the instruction to tease apart how much benefit they received from their training versus the instruction. Additional qualitative analysis will also help illuminate if students' language training played a role in how they answered and reasoned about the arithmetic problems, as was the case for student B01.

The higher performance on the temperature comparisons than the arithmetic problems makes sense given Case's (1996) framework. Based on his theory of number development, students need to understand the relations among number values before they can use this relation to add and subtract numbers. Further analysis of the ordering questions during the training session will clarify to what extent each student was able to map the directed magnitude language onto the integers.

The similar performance between second and fourth graders is interesting. On the one hand, the fourth graders performed higher than the second graders on the pretest. This is unsurprising as they had more time to be exposed to negative numbers and likely had more experience hearing about temperatures. However, second graders, even with less initial knowledge, were able to make the same gains as fourth graders! This suggests that the border established for when we introduce and/or teach negative integers should be reconsidered; second graders can make significant gains even with a relatively short intervention. If grade level doesn't matter, then having initial exposure to negative integers earlier could help students reach proficiency by the time the standards expect it.

Finally, second grade did better than fourth graders on problems, such as $4 - -5$, which are traditionally the hardest (Wheeler, Nesher, Bell, & Gattegno, 1981). It is possible that they were more successful at using the language to make sense of the problems; this is an area for further exploration. However, it suggests that younger students might benefit more from earlier exposure to the cognitively conflicting problems like these.

Acknowledgements

This research was supported by NSF CAREER award DRL-1350281.

References

- Bofferding, L. (2014). Understanding negative integers: Characterizing first-graders' mental models. *Journal for Research in Mathematics Education*, 45(2), 194-245.
- Bofferding, L. & Farmer, S. (2016, April). *Most and Least: Exploring Comparison Language with Integers*. Paper presented at the annual meeting of the American Education Research Association. Washington, DC.
- Case, R. (1996). Introduction: Reconceptualizing the nature of children's conceptual structures and their development in middle childhood. *Monographs of the Society for Research in Child Development*, 61(1-2), 1-26.
- DiBrienza, J., & Shevell, G. (1998). Number strings: Developing computational efficiency in a constructivist classroom. *The Constructivist*, 13(2), 21-25.
- Lampert, M., Beasley, H., Ghousseini, H., Kazemi, E., & Franke, M. (2010). Using designed instructional activities to enable novices to manage ambitious mathematics teaching. In M. K. Stein & L. Kucan (Eds.), *Instructional explanations in the disciplines* (pp. 129-141). New York, NY: Springer.
- Murray, P. L., & Mayer, R. E. (1988). Preschool children's judgments of number magnitude. *Journal of Educational Psychology*, 80(2), 206-209. doi: 10.1037/0022-0663.80.2.206
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author. Retrieved from <http://www.nctm.org/standards/content.aspx?id=26792>
- 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

- Palermo, D. S. (1973). More about less: A study of language comprehension. *Journal of Verbal Learning and Verbal Behavior*, 12(2), 211-221.
- Ryalls, B. O. (2000). Dimensional adjectives: Factors affecting children's ability to compare objects using novel words. *Journal of Experimental Child Psychology*, 76(1), 26–49. doi:10.1006/jecp.1999.2537
- Smith, L. B., Rattermann, M. J., & Sera, M. (1988). "Higher" and "lower": Comparative and categorical interpretations by children. *Cognitive Development*, 3(4), 341-357.
- Varma, S. & Schwartz, D. L. (2011). The mental representation of integers: An abstract-to-concrete shift in the understanding of mathematical concepts. *Cognition*, 121(3), 363-385.
- Wheeler, D., Neshier, P., Bell, A., & Gattegno, C. (1981). A research programme for mathematics education (I). *For the Learning of Mathematics*, 2(1), 27–29.