Title
Transitivity is Not Obvious: Probing Prerequisites for Learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors
Wiese, Eliane Stampfer
Patel, Rony
Olsen, Jennifer K
et al.

Publication Date
2015

Peer reviewed

Transitivity is Not Obvious: Probing Prerequisites for Learning

Eliane Stampfer Wiese (stampfer@cs.cmu.edu)
Rony Patel (rbpatel@andrew.cmu.edu)
Jennifer K. Olsen (jkolsen@cs.cmu.edu)
Kenneth R. Koedinger (krk@cs.cmu.edu)
Carnegie Mellon University, 5000 Forbes Avenue
Pittsburgh, PA 15232 USA

Abstract

Empirical results from a fraction addition task reveal a surprising gap in prior knowledge: difficulty applying the transitive property of equality in a symbolic context. 13 out of the 182 4^{th} and 5^{th} graders (7%) correctly applied the transitive property of equality to identify the sum of two fractions in a step-by-step worked example. This difficulty was robust to brief instruction on transitivity (after which performance rose to 11%). Students' demonstrated difficulty with transitivity is surprising, especially because common instructional techniques, such as worked examples, assume that the learner understands this concept and where it applies.

Keywords: conceptual understanding; fraction addition; mathematical equality

The Transitive Property of Equality: An Expert Blind Spot?

The transitive property of equality is fundamental to mathematics. It states that if a = b and b = c, then a = c. This property appears self-evident, perhaps explaining its absence from the common core state standards. However, even standards for the earliest grades rely on the application of this property. For example, the first grade standard "Add and subtract within 20" proposes four strategies for such problems, three of which use the transitive property (e.g., adding 8 + 6 with the 'making ten' strategy: "8 + 6 = 8 + 2 + 4 = 10 + 4 = 14"; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Left unstated in this example is that one may then conclude, by the transitive property, that 8 + 6 = 14.

Transitivity may seem so obvious to an expert that it becomes an "expert blind spot". Expert blind spots arise when the use of certain knowledge becomes so automatic that the expert does not realize it is being used. Blind spots may cause experts to incorrectly predict which tasks will be easy for novices, or to misdiagnose novices' difficulty (Asquith, Stephens, Knuth, & Alibali, 2007), and such difficulties may even be reinforced by common instructional designs (McNeil et al., 2006).

Transitivity and Fractions Concepts

The decision to investigate transitivity arose from the qualitative findings of a small-scale, exploratory pilot study with 6^{th} grade students, which we illustrate with two anecdotes. The experimenters (the first and second author) conducted the pilot at the students' school, where 7 students were pulled out of class, one at time, for about 20 minutes. The pilot was intended to assess the clarity of instruction and assessment materials for use in a future study. Initial items addressed fraction equivalence. When discussing equivalence with one participant, the participant agreed that 1/4 was equivalent to 5/20. The experimenters produced a 0-to-1 number line with 1/4 plotted, and the participant agreed that the mark showed 1/4. When asked where 5/20 would be plotted on the same number line, to our surprise, the participant indicated that it would fall to the left of 1/4. This response suggests that the student was not applying transitivity: the student agreed that 1/4 is equivalent to 5/20, and that 1/4 falls at a certain location on the number line, but did not combine these two facts to reason that 5/20 also falls at that location.

A discussion with another participant addressed how to add 1/4 and 1/5. The experimenters explained that the addends should first be converted to a common denominator. The participant converted both fractions to 20ths, and agreed that 1/4 was equivalent to 5/20, and 1/5 was equivalent to 4/20. After converting, the participant successfully added 4/20 and 5/20, yielding 9/20. However, when asked, "what is 1/4 plus 1/5?" the participant was unsure. The experimenters again verified that the participant agreed with the chain of steps: 1/4 was equivalent to 5/20, 1/5 was equivalent to 4/20, 1/4 + 1/5 was equivalent to 5/20 + 4/20, and the sum of 5/20 and 4/20 was 9/20. The participant agreed with each statement of equivalence, but still could not identify the sum of 1/4 and 1/5. Even when the experimenters explained that the sum was 9/20, the student still seemed a bit confused.

Together with similar areas of confusion demonstrated by other participants, the qualitative findings of the pilot suggested that middle school students had trouble applying the transitive property of equality, both when determining magnitude on a number line, and when reasoning about the equality of expressions and quantities in a multi-step problem. We hypothesize that difficulty applying transitivity in a fraction addition worked example is widespread, and is robust to brief instruction that points out the correct answer.

Transitivity Experiment

This transitivity experiment was part of the delayed post-test for a larger study that used within-class random assignment to compare three versions of an online fraction addition tutor. The main part of the larger study took place over three to four days, and included a pre-test, instruction and practice on fraction addition problems, and a post-test (all online). There were no significant effects of the larger study conditions on the transitivity experiment results (details in the results section). The delayed post-test took place 3 to 6 weeks after the main part of the study. The transitivity experiment consisted of the last items on the 31-item test. Each test item appeared sequentially, and students could not return to earlier questions. Students were not given any particular instruction around the transitivity items.

Materials

The transitivity experiment consisted of a pre-question, the solution to the pre-question, and then a post-question. In the pre-question, students were shown a solved fraction addition problem, including the sum, and were asked to enter the sum of the original addends (pre-question, Figure 1). After answering, students were shown a re-statement of the problem (Figure 2), and then pressed a button to see the answer, along with the instruction for their condition (Figure 4). The students were randomly assigned to one of four instructional conditions: (1) a conceptual text rule; (2) an example of procedural steps; (3) both; and (4) no instruction. The conceptual rule said “if you have three things and the first two are equal and the last two are equal, then all three are equal, so the first and the last are equal”. The example of procedural steps first highlighted that the right hand side of the first equation was the same as the left hand side of the second equation, then placed all three expressions on one line, with equal signs between them.

After seeing the solution to the pre-question (with accompanying explanation depending on the condition), students were given an isomorphic post-question. Both the pre- and post-questions used addends with denominators whose least common multiple (LCM) was smaller than their product, and the converted and sum fractions used the LCM as the denominator. These types of denominators help distinguish between students who are solving the problem from scratch (likely to use the product as the denominator) and students who are using transitivity to identify the sum.

The brief instruction is intended to clarify the assessment results, and not to provide in-depth teaching on the concept of transitivity. Problem statements usually do not explicitly include their solutions, and the instruction should reassure students that these are not a trick questions. Poor performance on the pre-item with high performance on the post-item would suggest that students understand transitivity, even if they need a quick reminder to use it. Poor performance on both the pre- and the post-item would suggest that students do not understand transitivity, at least in the context of fraction addition.

The delayed post-test also included 6 fraction addition items with unlike denominators, where neither denominator was a multiple of the other (Figure 3). Differences in performance between production and transitivity items would indicate that students are not solving the transitivity items from scratch.

Participants and Grade Level Standards

132 5th graders and 50 4th graders at a public school near Pittsburgh completed the transitivity experiment, which took place in school during the normal school day. The content of the larger fraction addition study aligns with the common core state standards for 4^{th} grade (finding equivalent fractions; same-denominator fraction addition) and 5^{th} grade (using equivalency to add fractions with unlike denominators; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). While the transitivity experiment involves unlike-denominator addition, a 5^{th} grade standard, the provision of

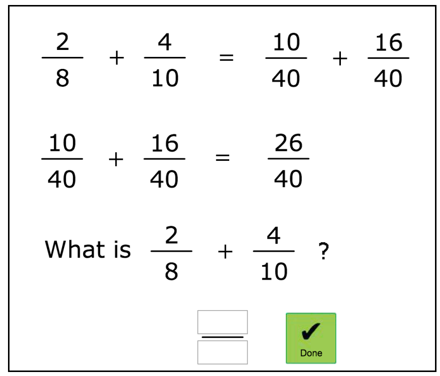

Figure 1: Pre-question, assessing if students realize the result of a multi-step problem is equal to the original problem expression.

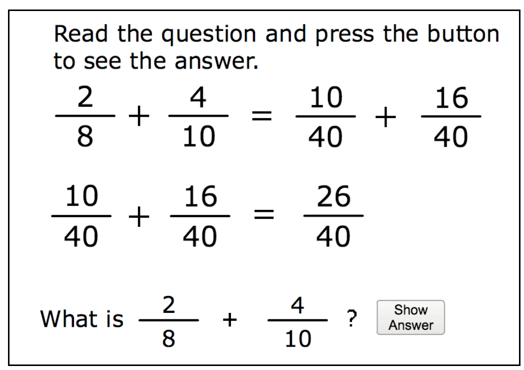

Figure 2: The first screen of the instruction. For the worked example conditions, the right-hand side of the first equation and left-hand side of the second equation were shown in purple to highlight that they are the same.

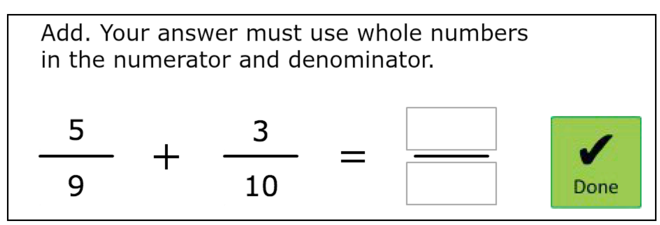

Figure 3: A fraction addition production item. A text field for optional scratch work was provided below (not shown)

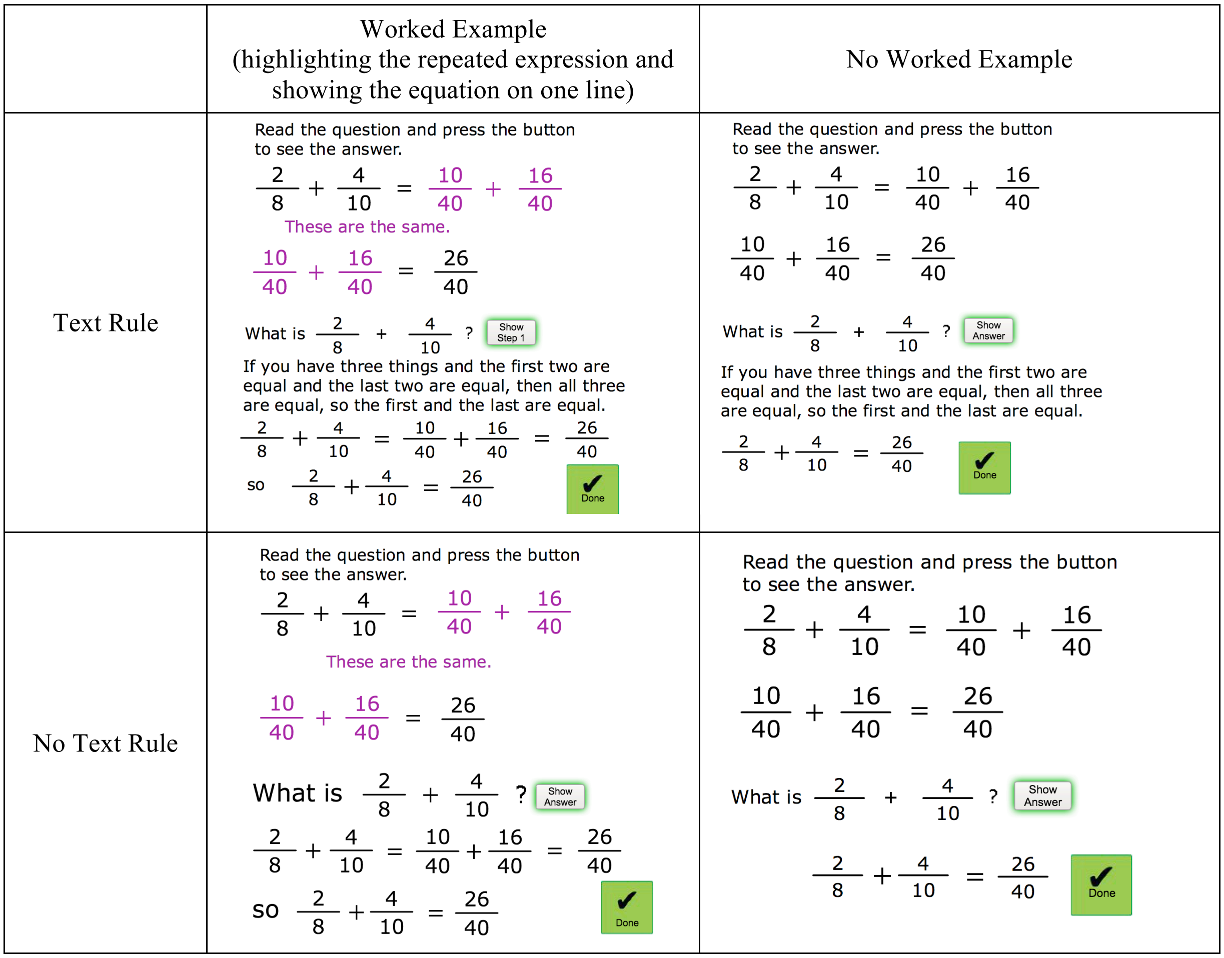

Figure 4: The second instruction screen for each condition. All conditions show the correct answer.

a worked example may move this content to a 4th grader's zone of proximal development. Therefore, although the qualitative pilot was with 6th graders, the common core state standards support the use of the 4th and 5th grade convenience sample from the larger study.

Results and Analysis

Table 1: Percentage of students per response category (absolute number of students in parenthesis, n = 182)

Answer Type	Pre-question	Post-question
Given Sum	.07 (13)	.11 (21)
Equivalent Sum	.09 (17)	.08 (13)
Incorrect Sum	.79 (143)	.79 (144)
Skipped	.05 (9)	.02 (4)

Table 1 shows student's answers, in four categories. *Given Sum:* the given sum provided by the second equation (26/40 in the pre-question). We expect that students who can apply the transitive property of equality will answer with the given sum. *Equivalent Sum:* a mathematically correct sum that is equivalent to the given sum (but is not the given sum). We expect that students who are adding the numbers from scratch will enter equivalent sums. While many responses are theoretically possible, in practice the only equivalent sums that students chose used denominators that were the product of the original denominators (e.g., 52/80 for the pre-question). *Incorrect Sum:* a fraction that is not equivalent to the correct sum. We expect that students attempting to add from scratch may do so incorrectly and will enter incorrect sums. *Skipped:* leaving either the numerator, denominator, or both blank. We expect that students who are confused by the question or are not motivated to answer will skip. 21 students (11%) entered the given sum on the post-question, compared with 13 students (7%) on the pre-question. However, collapsing response types into *given sum* and *other,* Fisher's exact test shows that the difference in response rate for *given sum* between pre- and post-question is not significant (p = .2).

Between-grade Comparisons All comparisons between the grades use Fishers exact test. The response rate for *given sum* on the pre-item was 4% for 4th graders (2 students) and 8% for 5th graders (11 students): this between-grade difference is not significant (p = .52). Students who answered with the given sum on the post-question but had not done so on the pre-question were categorized as 'transitivity learners.' 6% of 4th graders were transitivity learners (3 students) as were 4% of 5th graders (5 students). Again, this between-grade difference is not significant (*p* = .69). However, there were significant differences in the response rate for *equivalent sum,* both on the pre-item (*p* = .004) and the post-item (*p* = .021). None of the 4th graders entered equivalent sums, while 13% of 5th graders (17 students) did so on the pre-item and 10% (13 students) did so on the post-item. Between-grade differences for calculating the correct answer also held for the mean scores on production items: 17% correct for 5th grade, 0% correct for 4th. These results indicate that while 5th graders were better able to produce a correct sum by calculation, they did not outperform 4th graders on transitivity.

Stability of Responses All 13 students who entered the given sum on the pre-question also entered the given sum on the post-question. Additionally, 8 students who did not enter the given sum on the pre-question did enter the given sum on the post-question. However, this improvement-only pattern did not hold for the equivalent sum strategy. Students were about as likely to enter an incorrect sum on the first question and an equivalent sum on the second question as they were to do the reverse (3 and 4 students, respectively). For the most part, students gave the same answer type on the pre-question as they did on the post-question (158 students, 87%). Therefore, it is unlikely that students who answered with the given sum did so by randomly choosing between two equally likely options.

Response Times 13 students answered with the given sum on both the pre- and post-question. On average, they spent 42 seconds on the pre-question and 18 seconds on the post-question. A paired t-test showed that the difference in response times between the first and second question for students that responded with the given sum is significant ($p < .005$). 10 students answered with an equivalent sum on both the pre- and post-question. On average, they spent 37 seconds on the pre-question and 35 seconds on the post-question. A paired t-test showed that the difference in response times between the first and second question for students that responded with the equivalent sum is not significant ($p = .85$). These results indicate that students entering the given sum are not adding the original fractions from scratch and simplifying the sum.

Effect of Transitivity Instruction Type There were 8 transitivity learners (students who answered with the given sum on the post-question but not the pre-question): 1 in the example condition, 2 each in the answer-only and rule conditions, and 3 in the both condition. Fisher's exact test showed no significant difference in the number of transitivity learners among the four conditions (*p* = .95).

Average time spent on the instruction is given in Table 2. An ANOVA showed that the time spent on the instruction differed by instruction type (*p* = .02). Post-hoc tukey tests reveal that the *rule* and *both* instruction differ from each other ($p = .03$), with rule taking less time, and a marginal difference between *answer-only* and *both* ($p = .06$), with answer-only taking less time. If students are reading all of the instruction, *both* should take longest, *answer-only* should be shortest, and *example* and *rule* should be in between. The actual pattern of time taken is roughly consistent with this prediction, though perhaps students may not be reading all of the rule instruction.

Table 2: Mean time in seconds that students spent on the transitivity instruction, by instruction type

	Answer-only	Example	Rule	Both
Mean	13.1	16.7	12.5	19.6
Std. Deviation	11.9	12.5	8.2	15.9

Comparing Transitivity and Production When students answered with the given sum, how likely is it that they solved the problem from scratch instead of using transitivity? Of the 13 students who responded with the given sum on the transitive pre-question, 3 of those students solved all 6 production questions correctly, while the remaining 10 students solved 0 correctly. Of the 8 transitivity learners, 4 of them solved 5 or 6 production questions correctly, while the remaining 4 solved 0 correctly. These results suggest that many students who answered the transitivity items with the given sum were not able to add the original fractions. Conversely, students with a demonstrated ability to add fractions with unlike denominators did not automatically ignore the provided equations and solve the addition problem from scratch.16 students correctly answered all of the production items. 12 of them entered an equivalent sum on the pre-question, 3 entered the given sum, and 1 entered an incorrect sum. Together, these results show that transitivity is a separate skill from production: Students can score 0% on production items and still get transitivity items correct, and students can score 100% on production items and not recognize that the correct answer is already provided in a worked example.

The average score on production items was .12, and the average score on the two transitive items was .18 (counting both *given sum* and *equivalent sum* responses as correct). A paired t-test shows that this difference in scores is not significant (*p* = .3). This result indicates that overall, compared to production items, providing all of the steps to an unlike-denominator fraction addition problem *including the sum* did not significantly improve scores. However, students did not spend the same amount of time on both types of questions. On average, students spent 30 seconds on the transitive pre-question and 21 seconds on the production questions. A paired t-test shows that this difference is significant ($p < .005$). This result indicates that students were not simply ignoring the first two lines of equations and jumping to the addition question. Students took extra time on the transitivity pre-question compared to production items, likely because they were processing the equations. However, for many students the given equations appear to have been a distraction: compared with production items, the pre-question transitivity item took longer to solve and was not significantly more likely to be solved correctly.

While overall scores did not differ significantly between production items and the transitivity pre-question, certain errors occurred with different frequencies between the two question types. Table 3 shows response rates for mathematically correct answers and various errors between the production questions and transitivity questions. 60% of responses on the production questions used the incorrect strategy of obtaining the sum by adding both numerators and both denominators (add both error). 45% of responses on the transitivity items also demonstrated this error. A paired t-test on the add-both error rates shows that this difference is significant ($p < .005$). A small number of students entered sums that were equivalent to one of the addends (on production items) or were one of the two converted fractions shown in the given equations (on transitivity items). This error type was subdivided into *first converted* (entered sum is equivalent to the first addend or entered sum is the demonstrated first converted fraction) and *second converted* (likewise for the second addend or converted fraction). The rate of *first converted* errors was higher on transitivity items than production items (.027 vs. .0027). Collapsing response types into *first converted* and *other*, Fisher's exact test shows that the difference in frequency between production and transitivity items is significant ($p < .005$). While the rate of *second converted* is lower on transitivity items than production items (.005 vs. .014), Fisher's exact test shows that this difference is not significant ($p > .9$). Students who answered with *first converted* may have interpreted the first equation to mean that the left-hand side of the equation was equal to the first term of the right-hand side, instead of the entire expression on the right-hand side. Although this type of error is rare overall, it was 10 times more likely to occur on the transitivity items than production items. This suggests that for some students, difficulty in applying transitivity may stem from a misunderstanding of the equal sign.

Table 3: Error Rates on Production and Transitivity Items

Error	Production	Transitivity
Mathematically Correct	.12	.18
Add both	.60	.45
First converted	.0027	.027
Second converted	.014	.005
Skipped	.024	.035

No Effect of Larger Experimental Condition Fisher's exact test showed no significant differences between the three instructional conditions in the larger study, both for performance on the pre-question ($p = .9$), and for learning from the transitivity instruction ($p = .3$).

Discussion

Common instructional techniques for multi-step math problems, such as worked examples, assume that the learner understands transitivity. When learners themselves produce an answer to a multi-step problem, they appear to be demonstrating knowledge of transitivity. However, we are not aware of any previous work that has directly measured middle school students' understanding of transitivity in the context of mathematical symbols. This experiment shows that students' application of the transitive property of equality in a fraction addition context is very low: the given sum was entered for only 9% of answers across the two items. Strikingly, there was no significant difference in mathematically correct responses between production items and transitivity items – even though the transitivity items provided the converted fractions and the sum. Students' poor performance with transitivity was robust to brief instruction: though 8 students improved, this difference was not statistically significant, and there were no significant differences by instruction type. Average time spent on the instruction ranged from 13 seconds (rule) to 20 seconds (both), indicating that students were not completely ignoring the instruction. If students' poor performance on the pre-question was simply due to disbelief that the answer was really provided as part of the question, performance should have improved markedly after the brief instruction.

Validity of the Measures: The transitivity questions may over-estimate students' understanding, since students may arrive at the correct answer by adding in their heads using the least common multiple. However, most of the students who entered the given sum on the post-question did not answer a single production question correctly, suggesting that students were not using the same strategy on both. Further, students spent much longer on the transitive pre-question than the production questions (30 and 21 seconds, respectively), indicating that they were not simply ignoring the worked example. Conversely, the transitivity questions may under-estimate understanding, since those questions were the last test items and students may have been fatigued. However, the rate of mathematically correct responses to the transitivity items (18%) was not significantly different than the rate of such responses on production items (12%), which occurred earlier in the test. Further, if students were answering poorly due to fatigue, raw scores would not have improved between the pre- and post-item.

Why Did Students Fail to Apply Transitive Reasoning? On tasks involving physical objects, 5- to 6-year old children can apply transitivity more than 50% of the time (Andrews & Halford, 1998). Given that these 4^{th} and 5^{th} grade students must have some knowledge of transitivity, two possible explanations for their failure to apply it are cognitive load and misinterpretation of the equal sign. A cognitive load explanation would suggest that the task of interpreting the fraction symbols is so demanding that students do not have the cognitive headroom to apply transitivity. The test items in this experiment involved fractions with two-digit denominators and equations with operations on both sides. Items with reduced cognitive load could have whole numbers instead of fractions, or equivalence relations instead of operations. Comparisons of performance on items with different levels of cognitive load would indicate if students have difficulty with transitivity in all symbolic contexts, or just in contexts that are novel or complex. Indeed, on items asking if the sum of two positive numbers is greater than either addend alone, 5th graders' performance is lower for fractions than whole numbers (Wiese & Koedinger, 2014). We expect a similar pattern for transitivity.

Another explanation is that students misunderstand the meaning of the equal sign. Even students in grades 6-8 often misinterpret the equal sign to mean 'the total' or 'the answer' – interpretations that are not relational (McNeil et al., 2006). Indeed, one response from this study, answering with a fraction that is equivalent to the first addend, indicates a misunderstanding of the equal sign. Students may have thought the left-hand expression was equal to the first term that came after the equal sign rather than the entire right-hand expression. This error also points to possible difficulties in parsing and encoding equations with operations on both sides. McNeil et al. provide several assessments of students' understanding of the equal sign: verbal explanations, ratings of proposed explanations, and performance on mathematical equivalence tasks (McNeil & Alibali, 2000, 2005). Replicating these assessments in the context of fraction addition and equivalence tasks would illustrate if students' interpretations of the equal sign were affected by the fractions context. Assessing students' interpretation of the equal sign and application of transitivity in contexts with varying cognitive load would help tease apart these factors.

Finally, students may have performed poorly because of unfamiliarity with the materials. Though the intervention in the larger study involved symbolic fraction addition on a computer, it did not include worked examples. Specific instruction for students to read the worked example and to explain how the example relates to the question would help determine if any extraneous features of the item design impede students' performance.

Students' proficiency with fractions in middle school is a predictor for achievement in algebra (Siegler et al., 2012). However, the nature of this relationship between fractions and algebra is not well understood. Further investigations of the role of fundamental principles (such as transitivity) in both domains may shed light on this relationship.

Conclusions

The transitive property of equality is not obvious. However, this experiment does not show if recognition of the transitive property is equally difficult for middle school students in all contexts, or if the fractions context presents a particular difficulty. Future work should investigate middle school students' use of the transitive property with different types of numbers (i.e., whole numbers vs. fractions) and in different problem contexts (e.g., when studying an example vs. when producing an answer). These results further suggest that instruction that uses written or verbal examples may benefit from explicit assessments of how easily the target learners can identify the relationship between the problem and its solution.

Acknowledgments

This work is supported in part by Carnegie Mellon University's Program in Interdisciplinary Education Research (PIER) funded by grant number R305B090023 from the US Department of Education, and by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C100024 to WestEd. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Andrews, G., & Halford, G. S. (1998). Children's ability to make transitive inferences: The importance of premise integration and structural complexity. *Cognitive Development*, 13(4), 479–513.

Asquith, P., Stephens, A. C., Knuth, E. J., & Alibali, M. W. (2007). Middle School Mathematics Teachers' Knowledge of Students' Understanding of Core Algebraic Concepts: Equal Sign and Variable. *Mathematical Thinking and Learning*, 9(3), 249–272.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington, DC: Authors.

McNeil, N. M., & Alibali, M. W. (2000). Learning mathematics from procedural instruction: Externally imposed goals influence what is learned*. Journal of Educational Psychology*, 92(4), 734–744.

McNeil, N. M., & Alibali, M. W. (2005). Knowledge change as a function of mathematics experience: All contexts are not created equal. *Journal of Cognition and Development*, 6(2), 285–306.

McNeil, N. M., Grandau, L., Knuth, E. J., Alibali, M. W., Stephens, A. C., Hattikudur, S., & Krill, D. E. (2006). Middle-school students' understanding of the equal sign: The books they read can't help. *Cognition and Instruction*, 24(3), 367–385.

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., … Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science*, 23(7), 691–697.

Wiese, E. S., & Koedinger, K. R. (2014). Investigating Scaffolds for Sense Making in Fraction Addition and Comparison. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.