Authors' pre-publication copy.
Rau, M. A., Aleven, V., & Rummel, N. (2017). Supporting students in making sense of connections and in becoming perceptually fluent in making connections among multiple graphical representations. *Journal of Educational Psychology, 109*(3), 355–373. doi:10.1037/edu0000145

Supporting Students in Making Sense of Connections and in Becoming Perceptually Fluent in Making Connections among Multiple Graphical Representations

Martina A. Rau

Vincent Aleven

Nikol Rummel

Author Note

Martina A. Rau, Department of Educational Psychology, University of Wisconsin, Madison, USA;

Vincent Aleven, Human-Computer Interaction Institute, Carnegie Mellon University;

Nikol Rummel, Institute of Educational Research, Ruhr-Universität Bochum

Corresponding Author: Martina A. Rau, Department of Educational Psychology, University of Wisconsin, Madison, USA, marau@wisc.edu

Acknowledgements

This work was supported by the National Science Foundation, REESE-21851-1-1121307, by the IES R305A120734, and by the PSLC, funded by NSF award number SBE-0354420. We thank Richard Scheines, Ken Koedinger, Mitchell Nathan, Kathy Cramer, Jay Raspat, Michael Ringenberg, Brian Junker, Howard Seltman, Cassandra Studer, the students, teachers, and principals, the Tetrat, CTAT, and Datashop teams, especially Mike Komisin, Alida Skogsholm, and Joseph Ramsey.

Supporting Students in Making Sense of Connections and in Becoming Perceptually Fluent in Making Connections among Multiple Graphical Representations

Abstract

Prior research shows that multiple representations can enhance learning, provided that students make connections among them. We hypothesized that support for connection making is most effective in enhancing learning of domain knowledge if it helps students both in making sense of these connections and in becoming perceptually fluent in making connections. We tested this hypothesis in an experiment with 428 4th- and 5th-grade students who worked with different versions of an intelligent tutoring system for fractions learning. Results did not show main effects for sense-making or fluency-building support but an interaction effect, such that a combination of sense-making and fluency-building support is most effective in enhancing fractions knowledge. Causal path analysis of log data from the system shows that sense-making support enhances students' benefit from fluency-building support, but fluency-building support does not enhance their benefit from sense-making support. Our results suggest that both understanding of connections and perceptual fluency in connection making are critical aspects of learning of domain knowledge with multiple graphical representations. Findings from the causal path analysis lead to the testable prediction that instruction should provide sense-making support and fluency-building support for connection making.

Keywords

Multiple representations; connection making; intelligent tutoring systems; classroom evaluation; causal path analysis

1. Introduction

Instructional materials typically employ a variety of representations. For instance, students learning about fractions usually encounter the representations shown in Figure 1: circles, rectangles, and number lines. There is considerable evidence for benefits of multiple representations on students' learning (Ainsworth, 2006; de Jong et al., 1998; Eilam & Poyas, 2008). Multiple representations can enhance learning because they emphasize complementary conceptual aspects of the content (Larkin & Simon, 1987; Schnotz, 2005; Schnotz & Bannert, 2003). For example, the circle in Figure 1 depicts fractions as part of a whole circle, whereas the number line depicts fractions as a measure of length.

--- Insert Figure 1 about here ---

However, students' benefit from multiple representations depends on their ability to make connections among them (Ainsworth, 2006; Cook, Wiebe, & Carter, 2007; Taber, 2001). For example, learning of fractions requires an integration of the different concepts afforded by the representations in Figure 1 (National Mathematics Advisory Panel, 2008; Siegler et al., 2010). Therefore, students need to make connections among these representations. Yet, connection making is a difficult task (de Jong et al., 1998; Van Someren, Boshuizen, & de Jong, 1998) that students often fail to attempt spontaneously (Ainsworth, Bibby, & Wood, 2002; Authors, 2012a). At least two types of connection-making competencies play a role in students' learning. First, they need *understanding of connections*: the ability to map corresponding visual features of the GRs to one another (e.g., Ainsworth, 2006; Schnotz & Bannert, 2003; Seufert, 2003). For example, when working with the GRs in Figure 1, students may map the colored section in the circle to the number of sections between 0 and the dot in the number line, based on the rationale that both show the numerator of the fraction. Second, connection making involves the acquisition of *perceptual fluency:* learning to recognize visual patterns in GRs that correspond to domain-relevant concepts. For example, the student may learn to recognize that the GRs in Figure 1 show the same proportion of some unit.

Although prior research has yielded a number of effective interventions to support both types of connection-making competencies, this research has so far not investigated possible interactions among them. Our work addresses this gap by investigating whether combining support tailored to each type of connection-making competency enhances students' learning of fractions knowledge. We chose fractions as a domain for our research because—similar to many other STEM domains—instructional materials typically use multiple graphical representations (MGRs) that emphasize different concepts. Therefore, our research has the potential to generalize to other STEM domains. We conducted our research as part of regular classroom instruction in the context intelligent tutoring systems (Koedinger & Corbett, 2006), which are used in many classrooms across the United States and hence represent a realistic educational scenario. A further advantage of intelligent tutoring systems is that they allow for the use of interactive, virtual GRs while providing tutoring, which aligns with mathematics education research demonstrating advantages of virtual over physical GRs for fractions instruction (Moyer, Bolyard, & Spikell, 2002; Reimer & Moyer, 2005). For our experiment, we used the Fractions Tutor (Authors, 2013), which provides multiple virtual GRs and has been shown to yield significant learning about fractions knowledge among elementary-school students.

2. Motivation

2.1. Multiple Graphical Representations of Fractions

The mathematics education literature suggests that GRs fundamentally shape how students conceptualize fractions (Charalambous & Pitta-Pantazi, 2007; Cramer, Wyberg, &

Leavitt, 2008). Fractions are a notoriously complex domain (Charalambous & Pitta-Pantazi, 2007). Indeed, Behr, Lesh, Post, and Silver (1983) suggest at least six conceptual ways to interpret fractions: (1) parts of a whole, (2) decimals, (3) ratios, (4) quotient, (5) operators, and (6) measurements. GRs differ in their capacity to help students understand these concepts. For instance, area models (i.e., circles and rectangles) can illustrate part-whole concepts (e.g., one of four sections is shaded), ratio concepts (one section is shaded, three are unshaded), and quotient concepts (one whole divided by four) (Cramer et al., 2008). While circles are a type of area model in which the whole is inherent in the shape (i.e., a full circle; Cramer et al., 2008), rectangles do not have a standard shape but can be divided horizontally and vertically, which is helpful for illustrating quotient and operator interpretations. By contrast, linear models (e.g., number lines) are well suited to illustrate measurement and decimal concepts (Siegler et al., 2010).

Fractions instruction typically uses MGRs (Charalambous & Pitta-Pantazi, 2007; Kieren, 1993; Lamon, 1999; Martinie & Bay-Williams, 2003; Moss & Case, 1999; Thompson & Saldanha, 2003). Common curricula tend to start fractions instruction with area models (e.g., circles and rectangles) to introduce part-whole concepts of fractions and then work towards including other concepts, for instance by using number lines to illustrate measurement concepts (Behr et al., 1983; Kieren, 1976; Ohlsson, 1988). Failure to make connections among these different GRs may lead students to overly rely on one conceptual interpretation (Behr et al., 1983; Kieren, 1976; Ohlsson, 1988). This can cause misconceptions such as the “whole number bias”: the bias to treat fractions as composites of whole numbers (i.e., numerator and denominator), rather than as overall fraction values (Ni & Zhou, 2005). Indeed, Siegler and colleagues criticize early reliance on area models in fractions instruction for over-emphasizing part-whole concepts (Siegler et al., 2011, 2013). Instead, they recommend increased use of number line representations to emphasize measurement concepts. In line with this recommendation, educational practice guides emphasize advantages of number lines over other GRs (National Mathematics Advisory Panel, 2008; Siegler et al., 2010).

Given recent research on the potential privilege of number line representations over area models (Siegler et al., 2011, 2013), one may even argue that *unless* students make connections among GRs, they may learn better with number lines alone. Indeed, in our own prior research, we found that students benefited from MGRs only if they received instructional support to relate each GR to key fractions concepts (Authors, 2015). Without this support, students who worked with number lines alone showed higher learning gains than students who worked with MGRs.

In sum, students' benefit from MGRs depends on their ability to make connections among them. Yet, it remains an open question how best to support students in making such connections. We investigate this question in our current experiment. Because Siegler's suggestion that number lines alone may be more effective than MGRs is mainly rooted in concerns about failure to connect measurement concepts to part-whole concepts, our experiment focuses on connection making between the GRs typically used to emphasize these concepts: number lines and area representations (circle and rectangle).

2.2. Theoretical Framework

To address the question of how best to support students in making connections among MGRs, we draw on a recent theoretical framework that seeks to bridge cognitive science and educational research to educational practice: Koedinger and colleagues' (2012) Knowledge-Learning-Instruction framework (KLI; also see Koedinger et al., 2013). KLI offers the *alignment hypothesis:* that instructional interventions are most effective if they enhance learning processes that match the complexity of the to-be-learned competency. Hence, we use KLI to consider (1) the complexity of connection-making competencies that are important for domain expertise, (2) through which learning processes students acquire these competencies, and (3) which instructional interventions may match their complexity. As illustrated in Figure 2, these theoretical considerations lead to the hypothesis that combining support tailored to each type of connection-making competency enhances students' learning of fractions knowledge.

--- Insert Figure 2 about here ---

2.2.1. Connection-Making Competencies in Domain Expertise

The literature on expertise provides insights into how connection making among MGRs relates to domain expertise. Our review of this research suggests that two connection-making competencies play an important role in expertise (see Authors, in press, for an overview): *understanding* of connections (Ainsworth, 2006; Dreyfus & Dreyfus, 1986; Patel & Dexter, 2014; Richman et al., 1996) and *perceptual fluency* in connection making (Dreyfus & Dreyfus, 1986; Gibson, 1969, 2000; Pape & Tchoshanov, 2001; Richman et al., 1996). To analyze the complexity of these competencies, we draw on KLI's definition of a *knowledge component* as an "acquired unit of cognitive function […] that can be inferred from performance on a set of related tasks" (Koedinger et al., 2012, p. 764).

Understanding of connections among GRs means that a student can map visual features of one GR to those of a different GR because they show the same concept (Ainsworth, 2006; Charalambous & Pitta-Pantazi, 2007; Cramer, 2001; Kozma & Russell, 2005; Patel & Dexter, 2014). For example, consider a student who sees the GRs shown in Figure 1. The student may map the shaded section in the circle to the section between zero and the dot in the number line because both visual features depict the numerator, and he/she may relate the number of total sections in the circle to the sections between 0 and 1 in the number line because both features show the denominator. By reasoning about these connections, the student may understand the abstract principle that both GRs express fractions as portions of a unit, measured by partitioning the unit into equal sections. Under KLI, such reasoning involves learning of *complex* knowledge components because it requires that students learn a principle that applies in multiple situations (e.g., a proportion can be shown in multiple ways: circles, rectangles, number lines, etc.).

Perceptual fluency in making connections is the ability to quickly and effortlessly see holistic, corresponding visual patterns across different GRs. For example, a student should see “at a glance” that the circle and the number line show the same proportion of a unit. Perceptual fluency in connection making is related to domain expertise because it frees “cognitive head room” that allows students to reason about domain-relevant concepts (Gibson, 2000; Kellman & Massey, 2013; Richman et al., 1996). Under KLI, perceptual fluency involves learning of *simple* knowledge components because there is a one-to-one mapping between the GRs (e.g., circle and number line) and the visual pattern (e.g., proportion of unit covered).

2.2.2. Connection-Making Processes that Lead to Connection-Making Competencies

KLI moves beyond the analysis of knowledge components by relating them to the learning processes through which students acquire them. Students learn complex knowledge components via *sense-making processes*. These processes are verbally mediated because they involve explanations of principles of how GRs depict conceptually relevant information (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Gentner, 1983; Koedinger et al., 2012). They are explicit in that students have to willfully engage in them (Chi, de Leeuw, Chiu, & Lavancher, 1994; diSessa & Sherin, 2000). The literature on learning with representations often refers to sense-making processes as structure mapping processes (Gentner & Markman, 1997) because students map features of the representations to abstract concepts. Seufert (2003) suggests that structure mapping is one major process through which students integrate information from multiple representations into a coherent understanding of domain knowledge. diSessa's (2004) framework of meta-representational competence and research on representational flexibility (Acevedo Nistal, Van Dooren, & Verschaffel, 2013, 2015) suggest that sense-making processes are also involved in selecting appropriate GRs to solve domain-relevant problems.

By contrast, students learn simple knowledge components via *non-verbal inductive learning processes* (Koedinger et al., 2012; Richman et al., 1996) that they engage in when learning to categorize instances accurately and efficiently (Koedinger et al., 2012). These processes are often non-verbal because they do not require explicit reasoning (Kellman & Garrigan, 2009; Kellman & Massey, 2013). They are implicit because they typically happen unintentionally and unconsciously (Shanks, 2005) through experience with many instances (Gibson, 1969, 2000; Kellman & Massey, 2013; Richman et al., 1996). The literature also refers to inductive learning processes as perceptual learning and pattern recognition (Gibson, 1969, Goldstone & Barsalou, 1998; Kellman & Massey, 2013; Richman et al., 1996).

2.2.3. Instructional Interventions to Support Connection-Making Processes

According to KLI's *alignment hypothesis*, instructional interventions that enhance sense-making processes are most effective for complex knowledge components, whereas interventions that enhance inductive processes are most effective for simple knowledge components.

Supporting verbally mediated sense-making processes in connection making. KLI identifies principles that can guide the design of instructional activities that support sense-making processes (Koedinger et al., 2013). Here, we discuss two instructional activities that apply to the case of connection making: *explicitly comparing* multiple instances and providing *self-*

explanation prompts. Prior research has demonstrated how best to implement these principles into support for connection making. First, sense-making support is particularly effective if it prompts students to self-explain mappings between representations (Ainsworth & van Labeke, 2002; Bodemer & Faust, 2006; Seufert, 2003; Van der Meij & de Jong, 2011). Such prompts may be critical because students typically struggle in making sense of connections (Ainsworth et al., 2002), especially if they have low prior knowledge (Stern, Aprea, & Ebner, 2003). For example, Berthold and Renkl (2009) show that self-explanation prompts increase students' benefit from multiple representations. In their experiment, self-explanation prompts were implemented in the form of "why"-questions, in order to elicit self-explanations of principled knowledge. Self-explanation prompts are more effective if they ask students to self-explain specific connections than if they are open-ended (Berthold, Eysink, & Renkl, 2008; Van der Meij & de Jong, 2011).

Second, sense-making support typically asks students to use these mappings to compare how representations show analogous information or different, complementary information about the concepts they depict (Bodemer & Faust, 2006; Seufert, 2003; Seufert & Brünken, 2006; van der Meij & de Jong, 2006; Van Labeke & Ainsworth, 2002; Vreman-de Olde & De Jong, 2007). Although most implementations of sense-making support encourage students to compare representations, our review of prior research showed that there are two different, commonly used implementations. One common implementation of sense-making support in computer-based learning environments uses *linked representations*, where the student's manipulations of one GR are automatically reflected in the other GR (e.g., Ainsworth & van Labeke, 2002; Van der Meij & de Jong, 2006, 2011). Linked GRs allow students to explore intermediate steps, mistakes, and the final result in two or more GRs. This implementation aligns with KLI's cognitive dissonance principle, which states that presenting incorrect solutions may enhance sense-making processes (Koedinger et al., 2013).

A second common implementation uses analogous examples. These types of sense-making support typically provide step-by-step guidance for students map corresponding features across examples so as to extract their commonalities (e.g., Bodemer & Faust, 2006; Gutwill, Frederiksen, & White, 1999; Özgün-Koca, 2008). For example, Gutwill and colleagues (1999) found that asking students to map features of corresponding GRs to one another was effective in enhancing learning outcomes. Providing analogous examples aligns with KLI's worked example's principle (Koedinger et al., 2013).

Studies that compared the effects of sense-making support with linked representations and analogous examples yield contradictory findings. There is evidence in favor of linked representations (e.g., Van der Meij & de Jong, 2006, 2011), but there is also evidence in favor of analogous examples (e.g., Gutwill et al., 1999; Özgün-Koca, 2008). Hence, in the present experiment, we compare sense-making support with linked representations and with analogous examples, while incorporating self-explanation prompts in both.

Supporting non-verbal inductive refinement processes in connection making. KLI proposes that learning of simple knowledge components does not require that students engage in verbally mediated learning processes because there is nothing to explain. Evidence for this claim comes from studies showing that sense-making support is ineffective for simple knowledge components in perceptual learning (Schooler, Ohlsson, & Brooks, 1993; Schooler, Fiore, & Brandimonte, 1997) or grammar learning (Wylie, Koedinger, & Mitamura, 2009). KLI identifies a number of principles to guide the design of instructional activities that enhance non-verbal, implicit, inductive processes (Koedinger et al., 2013). Here, we discuss two principles that apply to perceptual fluency in connection making: immediate feedback and exposure to varied instances.

We note that the majority of connection-making support has focused on supporting sense-making processes rather than inductive processes. Yet, a new line of research yields a type of intervention that aligns with the KLI principles for inductive processes (Kellman & Massey, 2013; Kellman, Massey, & Son, 2009; Wise, Kubose, Chang, Russell, & Kellman, 2000). Kellman and colleagues developed interventions that provide fluency-building support for several science and mathematics topics (Kellman et al., 2009). These interventions ask students to rapidly classify representations over many short problems. In line with the KLI principle of immediate feedback, students receive correctness feedback on these problems. Further, the problems expose students to systematic variation, often in the form of contrasting cases, so that irrelevant features vary but relevant features remain constant across problems (Kellman & Massey, 2013). Studies in several domains (e.g., Kellman & Massey, 2013) show that fluency-building support leads to large and lasting gains in perceptual fluency that transfer to new instances and to learning gains on domain knowledge tests. Hence, in the present experiment, we investigate the effectiveness of fluency-building problems designed based on Kellman and colleagues' interventions.

2.3. Summary and Research Questions

In summary, KLI leads to the hypotheses we test in this paper, illustrated in Figure 2. We test the effects of sense-making problems that support verbally mediated sense-making processes (Figure 2, path 1) to enhance understanding of connections (Figure 2, path 5), and of fluency-building problems that support non-verbal inductive processes (Figure 2, path 2) to enhance perceptual fluency (Figure 2, path 6). We hypothesize that combining both types of connection-making support will enhance students' learning of fractions knowledge (Figure 2, paths 7 and 8).

This hypothesis remains untested because research on sense-making support and research fluency-building support are, to date, separate lines of research. In particular, prior research on sense-making support did not assess or manipulate students' perceptual fluency. Notably, most research on sense-making support involved connecting a GR to text-based representations. It seems reasonable to assume that students are fluent in reading (i.e., they have a high level of perceptual fluency in processing text). However, we do not know whether students in these studies had some level of perceptual fluency with the GR, and we do not know whether their level of prior perceptual fluency affected their benefit from sense-making support. Likewise, prior research on fluency-building support typically involved students who had already acquired conceptual understanding of the domain knowledge (e.g., Kellman et al., 2009), which is likely to involve understanding of connections. However, we do not know whether students' prior knowledge affected their benefit from fluency-building support.

We conducted a controlled classroom experiment that tested the following research questions and hypotheses:

Research question 1: Does connection-making support enhance students' learning gains?

Hypothesis 1.1: Students who receive sense-making problems that support connection making show higher learning gains of fractions knowledge than students who do not.

Hypothesis 1.2: Students who receive fluency-building problems show higher learning gains than students who do not.

Hypothesis 1.3: Students who receive a combination of sense-making and fluency-building problems show higher learning gains than students who receive either alone.

Research question 2: Are sense-making problems more effective if they include linked GRs or analogous examples?

This question was explorative, so we did not test specific hypotheses.

Research question 3: Does connection-making support enhance students' benefit from MGRs?

Hypothesis 3.1: Students who work with MGRs *without* connection-making support show higher learning gains than students who work with a single GR.

Hypothesis 3.2: Students who work with MGRs *with* connection-making support show higher learning gains than students who work with a single GR.

3. Classroom Experiment

3.1. Methods

3.1.1. Experimental design

--- Insert Table 1 about here ---

We randomly assigned individual students to work with one of several versions of the Fractions Tutor, which differed with respect to the types of connection-making problems they included. Our experiment had a 2x3 +1 design, summarized in Table 1. The two experimental factors were two types of connection-making problems: sense-making support and fluency-building support. The sense-making factor varied whether students received sense-making problems with linked representations (SL), sense-making problems with analogous examples (SE), or no sense-making problems. This factor was crossed with the fluency-building support factor, which varied whether students received fluency-building problems (F) or not. Students in the multiple-graphical-representations (MGR) condition received MGRs but no connection-making problems. Students in the single-graphical-representation (SGR) condition received only number lines and no connection-making problems.

3.1.2. Participants

599 4th- and 5th-grade students, aged 9-13 years, from five elementary schools (25 classes) in one school district in Pennsylvania, USA participated in the experiment. The school district was among the 10% highest ranked in reading and mathematics assessments of 500 Pennsylvania public school districts in the year of 2010/2011, with about 12% of students enrolled in free or reduced-price lunch programs, and 95% of the students being White. The school district volunteered to participate in this research.

3.1.3. Instructional Materials: The Fractions Tutor

We conducted the experiment in the context of the Fractions Tutor, an effective intelligent tutoring system designed for use in real classrooms (Authors, 2013). The Fractions Tutor supports learning through problem solving while providing immediate feedback and on-demand hints, both related to each problem step. The Fractions Tutor emphasizes conceptual learning by emphasizing principled understanding of fractions as proportions of a unit while students solve problems. The curriculum covers ten topics (see appendix in online supplemental material, Table 1A), covering about 10 hours of instruction. Students worked individually at their own pace. All conditions received 80 tutor problems: eight problems per topic, for ten fractions topics. For our experiment, we created different versions of the Fractions Tutor that varied what types of support for connection-making competencies it provides, detailed in the following. Consequently, the problems students encountered in the Fractions Tutor differed by condition, but we equated the number of problem-solving steps across conditions. Pilot-testing established that they took about the same time.

Single-graphical-representation (SGR) condition. Students in the SGR condition worked on number line problems only, eight per topic.

Multiple-graphical-representations (MGR) condition. Students in the MGR condition worked on eight individual-representation problems per topic. These problems involved only one GR per problem, but MGRs were used across problems, such that the students encountered each GR an equal number of times. Thus, the MGR condition received all three GRs, but no connection-making problems.

--- Insert Figure 3 about here ---

Figure 3 shows an example of an individual-representation problem. As students work through the steps of the problem, the Fractions Tutor provides feedback. The items shown in green are student entries with tutor feedback indicating that the value is correct, such as values entered in input boxes, selections from menus, and dots placed on an interactive number line. Students can also request a hint from the tutor on every step by clicking the brown button at the top right. Students interact with the GRs by using buttons to partition the GR into sections and by clicking to highlight sections in circles and rectangles or to place a dot on the number line. They also receive feedback on these interactions.

In the remaining five conditions, the first four problems for each topic were individual-representation problems. Students received the same number of individual-representation problems for each GR. The last four problems per topic were connection-making problems (i.e., sense-making problems with linked representations, sense-making problems with analogous examples, and/or fluency-building problems), corresponding to the student's experimental condition. Table 2 illustrates how sense-making problems and fluency-building problems were combined by contrasting three of the conditions.

--- Insert Table 2 about here ---

Sense-making with analogous examples (SE) condition. Students in the SE condition received four problems per problem in which they solved a part of the problem with one GR while being able to reference a set of worked-out steps for an analogous example that involved a different GR. These problems all share the same format, illustrated in Figure 4. Students were first given worked-out steps for a question with an area model (i.e., circle or rectangle; Figure 4-A, light green panel on the left). Next, the problem-solving part appeared on the right (Figure 4-B, light blue panel in the middle), with steps that were analogous to those in the example part. The problem-solving part always involved the number line. The key idea was that the analogous example uses the GR that is more familiar to students, given that—as mentioned above—fractions curricula tend to introduce fractions with area models. After completing the problem, students received self-explanation prompts to abstract a general principle from the two GRs (e.g., that both show equivalent fractions by re-partitioning the same amount; Figure 4-C, bottom). Self-explanation prompts were implemented in a fill-in-the blank format with drop-down menus on which students receive feedback. Similarly simple formats have been shown to be effective in prior research with intelligent tutoring systems or other educational technologies (Aleven & Koedinger, 2002; Atkinson, Renkl, & Merrill, 2003) and more effective than open-ended forms of self-explanation prompts (Gadgil, Nokes-Malach, & Chi, 2012; Johnson & Mayer, 2010; van der Meij & de Jong, 2011).

--- Insert Figure 4 about here ---

Sense-making with linked representations (SL) condition. Students in the SL condition received four problems per topic that included *support to make sense of connections with linked GRs* (Figure 5). Students interacted with a number line (Figure 5-A) to solve a problem, while an area model (i.e., a circle or a rectangle) updated automatically to mimic the same steps. Because students tend to be more familiar with area models than with number lines, linking was implemented such that the more familiar GR provided feedback on interactions with the less familiar GR. The SL problems included the same self-explanation prompts as SE problems (Figure 5-B).

--- Insert Figure 5 about here ---

Fluency-building (F) condition. Students in the F condition received four problems per topic that included *fluency-building support for connection making* (Figure 6). The fluency-building problems were designed based on Kellman and colleagues' (2009) interventions. Hence, they provided students with numerous short categorization problems. In the equivalent fractions topic, for instance, students sorted a variety of GRs using drag-and-drop (Figure 6). In alignment with Kellman and colleagues' interventions, fluency-building problems provided only correctness feedback. Students could request hints, but hint messages only provided general encouragement (e.g., "give it a try!"). Finally, the fluency-building problems encouraged visual problem-solving strategies. For example, in the equivalent fractions topic, students were instructed to visually judge equivalence rather than counting sections. To discourage counting strategies, we included examples with sections too small to count.

--- Insert Figure 6 about here ---

Combined sense-making and fluency-building conditions. Students in the sense-making with linked representations plus fluency-building (SL-F) condition also received four connection-making problems per topic: two SL problems followed by two F problems. Similarly, students in the sense-making with analogous examples plus fluency-building (SE-F) condition received two SE problems followed by two F problems. We decided to provide sense-making problems before fluency-building problems in each topic because understanding is expected before fluency in educational practice guides (e.g., National Council of Teachers of Mathematics, 2000, 2006).

3.1.4. Test instruments

Students took the tests three times: before they started working with the tutor (pretest), immediately after they finished working with the tutor (immediate posttest) and one week after the immediate posttest (delayed posttest). The delayed posttest was included so as to test whether students' knowledge is robust in that it lasts over time (Koedinger et al., 2012). We created three equivalent test forms, which included the same type of problems but with different numbers. We counterbalanced the order in which the different test forms were administered.

The tests targeted robust knowledge of fractions (i.e., with respect to domain knowledge, not connection-making knowledge) considering two knowledge types: procedural and conceptual knowledge. The conceptual scale included eight items that assessed students' principled understanding of fractions. The test items asked students to reconstruct the unit of a fraction, identify fractions from GRs, answer proportional reasoning questions, and complete written reasoning questions about fraction comparison tasks. The procedural scale included nine items that assessed students' ability to solve questions by applying algorithms. The test items asked students to find a fraction between two given fractions using GRs, finding equivalent fractions, addition, and subtraction. Both scales included multiple-choice and open-ended items. Half of the items in both test scales were reproduction and transfer items, respectively. Reproduction items were similar to individual-representation problems students had encountered during their work on the tutor. Transfer items were new relative to those covered in the tutor. The goal in including transfer items was to assess whether students' knowledge is robust in that it is transferred to unfamiliar problems (Barnett & Ceci, 2002). Example items for both tests can be found in the appendix in online supplemental material (Figures 1A, 2A). For questions that required multiple steps, partial credit was given for each correct step. The scores reported here are relative scores (i.e., ranging from 0 to 1). The theoretical structure of the test was based on a factor analysis with pretest data from the current experiment and was replicated with data from the immediate and delayed posttests. All test items were evaluated for their difficulty levels and discriminatory power using item-response-theory models. Taken together, the test items covered a range of difficulty levels. All items had good discriminatory power. Both scales had good reliability with Cronbach's Alpha of .70 for the conceptual scale and Cronbach's Alpha of .77 for the procedural scale.

3.1.5. Procedure

The study took place at the beginning of the 2011/2012 school year. Students accessed all materials online from their school's computer lab. They were instructed to work individually at their own pace with the Fractions Tutor. Classroom teachers led the sessions as they normally would during computer-lab hours; that is, they walked around to help individual students who needed assistance. They managed their classrooms in regular fashion; for instance, they told students to be quiet when they were chatting. Experimenters were present for the first two days of the experiment to ensure that the Fractions Tutor worked smoothly in the labs.

On day 1 of the study, students completed a 30-minute pretest. They then worked on the Fractions Tutor for about one hour per day for ten consecutive school days (i.e., two weeks, yielding about ten hours spent on the Fractions Tutor in total). On the last day, students completed a 30-minute posttest. One week later, students took a delayed posttest.

3.1.6. Analysis

Data in education research often has complex patterns of variance due to the fact that students are nested within classes (i.e., classes may account for a portion of the variance) and within schools (i.e., schools may account for a portion of the variance). Taking these sources of variance into account in statistical analyses allows to reduce the error variance statistical significance tests (Raudenbush & Bryk, 2002). Hierarchical linear models are a type of statistical model that allows accounting for such nested sources of variance (HLM; Raudenbush & Bryk, 2002).

We tested a number of variables, including teacher, school district, test form sequence, grade level, number of problems completed, total time spent with the tutor, random intercepts and slopes for classes and schools. We also tested whether including each level of the HLM increased model fit. The outcome of this selection procedure was the following four-level HLM. At level 1, we modeled performance on each of the tests for each student. At level 2, we accounted for differences between students. Level 3 models random differences between classes, and level 4 random differences between schools. Specifically, we used the following HLM:

$Y_{ijkl} = (((\mu + W_l) + V_{kl}) + \beta_2 * s_j + \beta_3 * f_j + \beta_4 * p_j + \beta_5 * s_j * p_j + \beta_6 * f_j * p_j + U_{jkl}) + \beta_1 * t_i + R_{ijkl}$

with

(level 1) $Y_{ijkl} = \varepsilon_{jkl} + \beta_1 * t_i + R_{ijkl}$

(level 2) $\varepsilon_{jkl} = \delta_{kl} + \beta_2 * se_j + \beta_3 * sl_j + \beta_4 * f_j + \beta_5 * p_j + \beta_6 * se_j * p_j + \beta_7 * sl_j * p_j + \beta_8 * f_j * p_j + U_{jkl}$

(level 3) $\delta_{kl} = \gamma_l + V_{kl}$

(level 4) $\gamma_l = \mu + W_l$

Table 3 provides an overview of the variables included in the HLM. Index *i* stands for test time (i.e., immediate and delayed posttest), *j* for the student, *k* for class, and *l* for the school.

The dependent variable Y_{ijkl} is student$_i$'s score on the dependent measures at test time t_i (i.e., immediate or delayed posttest), ε_{jkl} is the parameter for the intercept for student$_j$'s score, β_1 is the parameter for the effect of test time t_i, β_2 is the parameter for the effect of sense-making problems with analogous examples se_j, β_3 is the parameter for the effect of sense-making problems with linked representations sl_j, β_4 is the parameter for the effect of fluency-building problems f_j, β_5 is the parameter for the effect of student$_j$'s performance on the pretest p_j, β_6 is the parameter for an aptitude-treatment interaction between sense-making problems with analogous examples se_j and student$_j$'s performance on pretest p_j, β_7 is the parameter for an aptitude-treatment interaction between sense-making problems with linked representations sl_j and student$_j$'s performance on pretest p_j, β_8 is the parameter for an aptitude-treatment interaction between fluency-building problems f_j and student$_j$'s performance on pretest p_j, δ_{kl} is the parameter for the random intercept for class$_k$, γ_l is the parameter for the random intercept for school$_l$, and μ is the overall average. We ran this model in the SAS software package for mixed models.

--- Insert Table 3 about here ---

3.2. Results

We excluded students who did not complete all tests or did not complete the Fractions Tutor in the time allocated by their classroom teacher because they did not receive the full intervention and did not complete all topics that were tested in the posttests. The final sample included a total of $N = 428$ ($n = 61$ in the SGR condition, $n = 64$ in the MRG condition, $n = 52$ in the SL condition, $n = 59$ in the SE condition, $n = 73$ in the F condition, $n = 61$ in the SL-F condition, $n = 59$ in the SE-F condition). The number of students who were excluded from the analysis did not differ significantly between conditions, χ^2 (6, $N = 169$) = 4.34, $p > .10$. Excluded students had significantly lower pretest scores on the conceptual knowledge test, $F(1,594)$ =

6.73, $p < .05$, and on the procedural knowledge test, $F(1,594) = 5.60$, $p < .05$, but there were no differences between conditions (Fs < 1). Students' lower prior knowledge may explain why they took longer in working with the Fractions Tutor and, hence, did not finish in the allocated time.

Table 4 shows the means and standard deviations for the conceptual and procedural scales by test time and condition. Table 5 shows the total amount of time spent on tutor problems by condition. To verify that time spent did not differ between conditions, we used the same HLM as described above. There were no significant effects of sense-making support, fluency-building support, nor a significant interaction among these factors on time spent (Fs < 1).

--- Insert Table 4 about here ---

--- Insert Table 5 about here ---

3.2.1. Learning gains

In learning experiments in real educational settings, any difference between conditions needs to be interpreted relative to pretest-to-posttest learning gains (Lipsey et al., 2012). Thus, we first verified whether students learned from the Fractions Tutor. To do so, we used a modified version of the HLM described above on all seven conditions, using pretest scores as a repeated, dependent measure rather than as a covariate (the SAS-code can be found in the appendix in online supplemental material, Figure 3A). Students performed significantly better on conceptual knowledge at the immediate posttest ($p < .0001$, $d = .40$), and at the delayed posttest ($p < .0001$, $d = .60$), compared to the pretest. Students performed significantly better on procedural knowledge at the immediate ($p < .0001$, $d = .20$) and at the delayed posttest ($p < .0001$, $d = .24$), compared to the pretest.

3.2.2. Effects of connection-making support

To investigate whether a combination of sense-making problems and fluency-building problems leads to higher learning gains than either type of problem alone, we applied the HLM described above to the 2x3 design (i.e., without the SGR condition; the SAS-code can be found in the appendix in online supplemental material, Figure 4A). The parameter estimates can be found in the appendix in online supplemental material (Tables 2A for random intercepts, 3A for fixed effects in the conceptual knowledge model, Table 4A for fixed effects in the procedural knowledge model). There were no main effects of sense-making problems (hypothesis 1.1) or fluency-building problems (hypothesis 1.2) on conceptual knowledge or on procedural knowledge (Fs < 1). There were no significant interactions of sense-making problems or fluency-building problems with pretest performance. There was no significant interaction on procedural knowledge (Fs < 1). However, there was a significant interaction between sense-making problems and fluency-building problems on conceptual knowledge, $F(2, 343) = 4.11$, $p = .017$, $\eta^2 = .03$, such that students who received both types of problems performed best on the conceptual posttests. To gain further insights into this interaction effect, we turn to research question 2: are sense-making problems more effective if they include linked GRs or analogous examples? We examined simple effects of the sense-making factor for the conditions with fluency-building problems (i.e., SL-F, SE-F, F conditions) and without fluency-building problems (i.e., SL, SE, MGR conditions). On conceptual knowledge, there was a significant effect of sense-making problems among the conditions with fluency-building problems, $F(2, 343) = 4.34$, $p = .014$, $\eta^2 = .07$, such that the SE-F condition significantly outperformed the F condition, $t(341) = 2.82$, $p = .005$, $d = .32$, and the SL-F condition, $t(342) = 2.20$, $p = .05$, $d = .26$. The difference between the SE-F condition and the F condition was not significant ($t < 1$). The effect of sense-making problems was not significant for the conditions without fluency-building problems ($F < 1$), and consequently, none of the post-hoc comparisons were significant.

To investigate whether MGRs are more effective than an SGR (hypotheses 3.1 and 3.2), we applied a modified version of the HLM described above to the SGR, MGR, and SE-F condition (i.e., the most successful connection-making condition; the SAS-code can be found in the appendix in online supplemental material, Figure 4A). There were no significant differences between the MGR condition and the SGR condition (*p*s > .10) (hypothesis 3.1). The SE-F condition significantly outperformed the SGR condition on conceptual knowledge, $t(115) = 2.41$, $p = .016$, $d = .27$, but not on procedural knowledge ($t < 1$) (hypothesis 3.2).

3.3. Discussion

With respect to research question 1 (does connection-making support enhance students' learning gains?), our results do not support hypotheses 1.1 or 1.2, that problems that work on sense-making or working on fluency-building problems would enhance robust fractions knowledge, respectively. However, our results support hypothesis 1.3 for conceptual knowledge: working on a combination of sense-making problems and fluency-building problems was effective. Somewhat to our surprise, neither type of connection-making support alone, but *only* the combination of both was effective. With respect to research question 3 (does connection-making support enhance students' benefit from MGRs?), our results stand in contrast to hypothesis 3.1 but support hypothesis 3.2. Comparisons to the SGR condition show that students did not benefit from working with MGRs, unless they received a *combination* of sense-making and fluency-building support.

We did not find significant effects on procedural knowledge. It may be that students' conceptual knowledge benefits from connection making because each representation provides a different conceptual view on what fractions are, whereas procedural knowledge may rely more on experience with algorithmic operations tasks rather than on conceptual understanding.

With respect to our exploratory research question 2, whether problems that help students make sense of connections are more effective if they include linked GRs or analogous examples, we find that analogous examples lead to higher learning gains on a test of robust fractions knowledge than linked GRs.

4. Causal Path Analysis Modeling

The experiment showed that only the *combination* of sense-making problems and fluency-building problems was effective in enhancing students' learning of domain knowledge. This finding leads to open questions about *how* sense-making processes and inductive refinement processes interact (Figure 2, paths 3 and 4). Hence, we seek to better understand the nature of this interaction through an additional data source–the tutor log data as an indicator of problem-solving performance–using causal path analysis modeling. The logs provide a detailed record of students' interactions with the Fractions Tutor at the "transaction" level (i.e., attempts at steps, hint requests, etc.). Given that sense-making problems with analogous examples were more effective than those with linked GRs, we focused on the SE conditions in this analysis.

4.1. Hypotheses

We investigate two possible mechanisms by which sense-making problems and fluency-building problems might interact. One mechanism may be that working on fluency-building problems enhances students' benefit from sense-making problems (Figure 2, path 3; we will refer to this as the *fluency hypothesis*). According to the *fluency hypothesis*, perceptually fluent students may benefit from increased cognitive capacity during subsequent learning tasks (Kellman et al., 2009; Koedinger et al., 2012). Therefore, they should show better performance on sense-making problems. We contrast the fluency hypothesis to the *practice hypothesis* that receiving more practice on sense-making problems leads to better performance on sense-making problems. The SE condition has practice on four sense-making problems, per problems whereas the SE-F condition has practice on only two sense-making problems. Therefore, the practice hypothesis predicts that the SE condition should show better performance on sense-making problems than the SE-F condition. To see the effect of having practice with fluency-building problems on students' performance on sense-making problems, we compare the SE condition to the SE-F condition. In the SE condition, problems P5, P6, P7, and P8 were sense-making problems (for each of the ten topics, see Table 2). In the SE-F condition, only problems P5 and P6 were sense-making problems (for each of the ten topics). Hence, when comparing the SE and SE-F conditions, problems P5 and P6 of each topic serve as the basis for the comparison (bold-underlined problems in Table 2).

Another mechanism may be that working on sense-making problems enhances students' benefit from fluency-building problems (Figure 2, path 4; *sense-making hypothesis*). Prior research shows that students have difficulties in making sense of connections at a conceptual level and typically do not make connections spontaneously (Ainsworth et al., 2002; Authors, 2012a). Therefore, the sense-making hypothesis predicts that students may not be able to discover what features of the GRs depict meaningful information while working on fluency-building problems, which may lead to inefficient learning strategies (e.g., trial-and-error) that can impede their benefit from fluency-building problems. In particular, the visual features that denote fractions may not be easy to detect, and can perhaps not be learned in a purely inductive manner. Therefore, sense-making support could increase students' performance on fluency-building problems. We contrast the sense-making hypothesis to the practice hypothesis that receiving more practice on fluency-building problems leads to better performance on fluency-building problems. The F condition has practice on four fluency-building problems per problems, whereas the SE-F condition has practice on only two fluency-building problems. Therefore, the practice hypothesis predicts that the F condition should show better performance on fluency-building problems than the SE-F condition. To investigate the effect of having practiced on sense-making problems on students' performance on fluency-building problems, we compare the F condition to the SE-F condition. In the F condition, problems P5, P6, P7, and P8 were fluency-building problems (for each of the ten topics, see Table 2). In the SE-F condition, only problems P7 and P8 were sense-making problems (for each of the ten topics). Hence, when comparing the F and SE-F conditions, problems P7 and P8 for each topic serve as the basis for the comparison (bold-italicized problems in Table 2).

In testing the fluency hypothesis and the sense-making hypothesis, we allow for the possibility that they are not mutually exclusive.

4.2. Methods

To investigate these hypotheses, we use causal path analysis, which provides a unified framework to test mediation hypotheses, estimate total effects, and separate direct from indirect effects in a coherent statistical model (Bollen & Pearl, 2013; Chickering, 2002; Spirtes et al., 2000). We constructed causal path analysis models that correspond to the fluency hypothesis and to the sense-making hypothesis, respectively.

Because we selected the SE and SE-F conditions for the fluency hypothesis model and the F and SE-F conditions for the sense-making hypothesis model, 190 students were included in the analysis (n = 59 in the SE condition, n = 73 in the F condition, and n = 58 in the SE-F condition).We operationalized performance on the tutor problems as error rates: making fewer errors while solving a tutor problem indicates higher problem-solving performance. Rather than using the overall error rate, we classified errors based on the detailed knowledge components to which they relate. For the fluency hypothesis model, we computed the error rate for each knowledge component across the sense-making problems P5 and P6 for all ten topics (bold-underlined problems in Table 2). For the sense-making hypothesis model, we computed the error rate for each knowledge component across the fluency-building problems P7 and P8 for all ten topics (bold-italicized problems in Table 2). Altogether, the knowledge component model yielded twelve error types that students could make on sense-making problems, and eleven error types that students could make on fluency-building problems, summarized in Tables 7 and 8. Next, included only those error types in the causal path analysis model that (1) were significant predictors of performance on the conceptual posttest, while controlling for pretest, and (2) significantly differed between conditions (i.e., the italicized error types in Tables 7 and 8).

--- Insert Table 7 about here ---

--- Insert Table 8 about here ---

We constructed the causal path analysis models using an automatic algorithm that searches for models that are theoretically plausible and consistent with the data; namely, the Tetrad IV program's[1] GES algorithm. Tetrad IV allows us to specify assumptions that constrain the space of models searched (Chickering, 2002; Spirtes et al., 2000) and to find the model with the best model fit among models that are theoretically tenable and compatible with the experimental design (Spirtes et al., 2000). *Independent variables* in the causal path analysis were sense-making support and fluency-building support. *Dependent variables* were students'

[1] Tetrad, freely available at www.phil.cmu.edu/projects/tetrad, contains a causal model simulator, estimator, and over 20 model search algorithms, many of which are described and proved asymptotically reliable in (Spirtes, Glymour, & Scheines, 2000).

performance on the conceptual pretest, immediate and delayed posttest. *Mediators* were error types students made on the sense-making problems for the fluency hypothesis model, and error types students made on the fluency-building problems in the sense-making hypothesis model.

When conducting a model search, we can narrow the search space based on the knowledge we have about the nature of our data (Spirtes et al., 2000). We assumed that the experimental conditions are exogenous and causally independent, that the pretest was not influenced by the conditions, that the pretest is an exogenous variable and causally independent of the conditions. Furthermore, we assume that the mediators are prior to the immediate posttest and the delayed posttest, and that the immediate posttest is prior to the delayed posttest. The search space is defined by the fully saturated model for each hypothesis because it contains all possible edges (or "effects") compatible with these assumptions and with the experimental design.

We had Tetrad search among models that had all, none, or a subset of the edges in the fully saturated model. In the model search, each edge is automatically evaluated as to whether including it yields a better model fit than not including it, and whether it represents a statistically significant effect. Figure 7 (left) illustrates the fully saturated model for the fluency hypothesis (which includes only performance variables related to sense making as possible mediators). Figure 7 (right) illustrates the fully saturated model for the sense-making hypothesis (which includes only performance variable related to perceptual fluency as possible mediators). Thus, Figure 7 illustrates that, even with our assumptions, the search space contains at least 2^{15} (over 32 thousand) distinct path models that are plausible tests for the sense-making hypothesis, and 2^{20} (over 1 million) for the fluency hypothesis. The outcomes of the model search are two causal path analysis models, one corresponding to the fluency hypothesis, one corresponding to the sense-making hypothesis, each consistent with the data and hence allowing us to trust the parameters of the model.

--- Insert Figure 7 about here ---

4.3. Results

--- Insert Figure 8 about here ---

To test the fluency hypothesis, we inspect the model shown in Figure 8, which is the best-fitting model Tetrad IV found for the fluency hypothesis. The model fits the data well ($\chi^2 = 8.32$, df = 5, $p = .14$; CFI = 0.9943; RMSEA = 0.0808)[2]. The standardized coefficients and their standard errors, the significance tests for each effect, and the implied covariance matrices for the model are provided in the appendix in online supplemental material (Tables 5A, 6A, 7A). Figure 8 shows unstandardized coefficients, which are easier to interpret with respect to the effects of number of errors students made. Further, because scores on all tests range between 0 and 1, the effects on the posttests are easy to compare even though coefficients are unstandardized. Recall that this model compares the SE and SE-F conditions based on errors students made on the sense-making problems. Further recall that, according to the fluency hypothesis, we expect that practice on fluency-building problems reduces error rates on sense-making problems, and that error rates on sense-making problems mediates the effect of condition on the posttests. Finally, recall that the alternative practice hypothesis suggests that, because students in the SE-F condition have less practice on sense-making problems, they should show higher rates of sense-making errors. The model in Figure 8 shows that students in the SE-F condition, compared to the

[2] The usual logic of hypothesis testing is inverted in path analysis. The *p*-value reflects the probability of seeing as much or more deviation between the covariance matrix implied by the estimated model and the observed covariance matrix, conditional on the null hypothesis that the model that we estimated was the true model. Thus, a low *p*-value means the *model* can be rejected, and a high *p*-value means it cannot. Conventional thresholds are .05 or .01, but like other alpha values, this is somewhat arbitrary. The *p*-value should be higher at low sample sizes and lowered as the sample size increases, but the rate is a function of several factors, and generally unknown.

SE condition, made *more* selfExplanationErrors (i.e., the average student in the SE-F condition made 5.662 more errors in answering self-explanation prompts than the average student in the SE condition, and for each of these errors, the student looses .005 points on the final posttest) and *more* place1Errors (i.e., errors in finding 1 on a number line). Both decreased learning gains. Thus, students' performance on sense-making problems mediated a negative effect of fluency-building support on students' posttest performance. This negative mediation effect is in line with the alternative hypothesis that practice alone explains performance on sense-making problems. In addition, in line with the overall finding of the experiment, Figure 8 shows that fluency-building support had a direct positive effect on posttest performance, which was stronger than the negative mediation effects. That is, the direct path of .116 is larger than the sum of the mediating paths (-.005 * 5.662 + -.012 * .166 * 5.662).

--- Insert Figure 9 about here---

To test the sense-making hypothesis, we inspect the model in Figure 9, which shows the best-fitting model for the sense-making hypothesis. This model fits the data reasonably well (χ^2 = 16.10, df = 6, *p* = .013; CFI = 0.9822; RMSEA = 0.1338)[3]. The standardized coefficients and their standard errors, the significance tests for each effect, and the implied covariance matrices for the model are provided in the appendix in online supplemental material (Tables 5A, 6A, 7A). Figure 9 shows the unstandardized coefficients. Recall that this model compares the F and SE-F conditions based on errors students made on the fluency-building problems. Further recall that, according to the sense-making hypothesis, we expect that practice on sense-making problems leads to a lower rate of errors on the fluency-building problems, which in turn mediates the effect

[3] Ibid. It is worth noting that this model asserts that any effect the SE-F condition (compared to the F condition) has on the post-test or delayed post-test is entirely mediated by the three variables measuring error rates. Thus, it makes a bold and easily falsifiable prediction that is tested by this model.

of condition on the posttests. Finally, recall that the alternative practice hypothesis suggests that, because students in the SE-F condition have less practice on fluency-building problems, they should show higher error rates on fluency-building problems.The model in Figure 9 shows that students in the SE-F condition made more nameCircleMixed errors (i.e., errors in identifying the fraction depicted by a circle) but fewer improperMixedErrors (i.e., errors in identifying an improper fraction) and fewer equivalence errors (i.e., errors in identifying equivalent fractions) than students in the F condition. Students who made fewer nameCircleMixedErrors also made more subtractionMixedErrors (i.e., errors in finding the difference between two given fractions) and improperMixedErrors, which decreased performance in the conceptual posttest. Thus, performance on fluency-building problems mediated the positive effect of sense-making support on the conceptual posttest. There were no additional direct effects of sense-making support on posttest, so that students' higher performance on fluency-building problems fully mediated the positive effect of sense-making support on learning gains.

4.4. Discussion

The results from the causal path analysis are consistent with the sense-making hypothesis but stand in contrast to the fluency hypothesis: we did not find evidence that working on fluency-building problems *helps* students benefit from sense-making problems, but that fluency-building problems *decrease* their performance on sense-making problems. Thus, the mediation effect shown in Figure 8 suggests that receiving fluency-building problems comes at the cost of lower performance on sense-making problems: students tend to make more selfExplanationErrors and more place1Errors. Recall that that students in the SE condition work on twice as many sense-making problems than the SE-F condition, so they receive more practice on these problems compared to the SE-F condition (see Table 2). Hence, the practice hypothesis predicts that they perform somewhat worse on those problems, simply because they have less practice. The model in Figure 8 is in line with the practice hypothesis. Furthermore, the model in Figure 8 puts the performance on sense-making problems in relation to learning gains: higher performance on sense-making problems is associated with higher learning benefit from sense-making problems. Yet, since we do not find evidence that fluency-building problems help students learn from sense-making problems, our results do not support the fluency hypothesis.

By contrast, the results from the causal path analysis models are in line with the sense-making hypothesis: working on sense-making problems helps students learn from fluency-building problems. The model in Figure 9 demonstrates that, although students who receive sense-making problems make more nameCircleMixedErrors, they make fewer equivalenceErrors and improperMixedErrors while working on fluency-building problems. The reduction of equivalenceErrors and improperMixedErrors mediates the effect of sense-making support on learning gains. NameCircleMixedErrors are confined to an early topic in the Fractions Tutor, whereas equivalenceErrors and improperMixedErrors occur later in the Fractions Tutor. The results therefore suggest that working on sense-making problems reduces errors later during the learning phase, which leads to higher learning gains. This finding is particularly interesting because it indicates that having worked on sense-making problems leads to higher performance on fluency-building problems, *even though* students in the F condition had more practice opportunities on fluency-building problems (practice hypothesis). Thus, it seems that sense-making problems prepare students to benefit from subsequent fluency-building problems—even more so than practice with fluency-building problems does.

5. General Discussion

Our experiment investigated how best to support students in making connections among MGRs. Our results support our hypothesis that a combination of sense-making and fluency-building support is most effective with respect to learning of conceptual knowledge. Surprisingly, we found that *only* the combination of sense-making problems and fluency-building problems is effective; taken alone, neither sense-making problems nor fluency-building problems were effective. By establishing that sense-making problems and fluency-building problems interact, this finding extends prior research that has—to the best of our knowledge—exclusively focused on either sense-making support (e.g., Seufert, 2003; Bodemer & Faust, 2006; van der Meij & de Jong, 2006), or on fluency-building support (e.g., Kellman et al., 2009). As argued above, students in prior research on sense-making support may have had some level of perceptual fluency in interpreting the representations used in these studies (i.e., mostly text-based and numerical representations). Likewise, students in prior research on fluency-building support likely had, to some extent, understanding of connections because they had typically received prior instruction on the domain knowledge. Our finding that both types of support are necessary does not necessarily contradict prior research. Rather, our findings extend it by indicating that the aspects that were held constant across conditions in prior research may be an important prerequisite to the effectiveness of either type of support. At a practical level, our results suggest that standard sense-making support should take into account students' level of perceptual fluency. Instructors may need to ensure that students are indeed perceptually fluent in making connections, in which case sense-making support alone could be effective (although this hypothesis has not been tested), or they might need to combine sense-making support with fluency-building support (as in our experiment).

It is also interesting to reflect on the fact that we did not find evidence that MGRs without connection-making support lead to higher learning gains than a single GR that is considered the "superior" GR by some researchers: the number line (National Mathematics Advisory Panel, 2008; Siegler et al., 2010). We found that MGRs were more effective than a single GR *only if* students received a combination of sense-making and fluency-building support. This finding is in line with our own prior research (Authors, 2012b), which shows that MGRs are not always effective in enhancing fractions learning. It is also in line with experiments in other domains that failed to show a benefit of MGRs over learning with a single GR (e.g., Berthold & Renkl, 2009; Corradi, Elen, & Clareboug, 2012). MGRs are commonly used in instruction because they emphasize multiple conceptual perspectives. Our results support this practice but also caution that integrating these conceptual perspectives into their domain knowledge is a difficult task for students. To support them in doing so, instruction may need to provide a combination of sense-making support and fluency-building support.

The causal path analysis models provide additional insights into the mechanisms underlying this finding. We found that sense-making problems enhance students' benefit from fluency-building problems by reducing the number of certain types of errors students make on fluency-building problems. Hence, understanding of connections seems to provide the foundation for inductive processes that students engage in when working on fluency-building problems. Our findings do not support the reverse conclusion: we have no evidence that fluency-building problems enhance students' benefit from sense-making problems. In contrast, we found that more practice on sense-making problems yields better performance on sense-making problems, as expected purely based on practice effects. Thus, it seems that, even if there are benefits of additional cognitive headroom as a result of perceptual fluency, they do not outweigh the advantages of practice effects on the same type of problem.

Lipsey and colleagues (2012) suggest that effect sizes of interventions obtained in real classrooms must be interpreted in relation to pretest-to-posttest changes. Ranging between d = .20 and d = .60 resulting from a ten-hour long intervention, effect sizes of learning gains are of small to medium size. According to Hattie's (2012) meta-analysis of educational interventions in realistic settings, the average effect size of interventions are d = .40 per year on student achievement (e.g., p. 16, p. 240 in Hattie, 2012). Thus, our experiment shows learning gains that compare favorably to those obtained in other studies. A similar argument can be made when interpreting the effect sizes for the between-condition effects. The advantage of receiving a combination of sense-making and fluency-building support compared to working with only the number line representation had an effect size of d = .27. Thus, comparing this difference to the learning gain of d = .40 on the conceptual knowledge test, the benefit of combining sense-making problems and fluency-building problems when providing students with MGRs seems meaningful.

It is important to note a number of limitations of this research. First, we excluded students who did not finish their work with the Fractions Tutor because they did not receive full exposure to the experimental intervention and because the posttests assessed knowledge targeted in all topics of the curriculum. However, this decision led to excluding many students, and these students had lower pretest scores than students who were included in the analysis. Because students were randomly assigned to conditions and because the number of excluded students did not differ by conditions, this exclusion does not undermine our overall conclusions, but implies that future research should test that our findings generalize to lower-performing students. We also note that the school population was mostly White and included only a small portion of students from low-income families. Although we cannot think of a reason why students from more diverse backgrounds would not benefit from a combination of sense-making and fluency-building support, future research should empirically verify this prediction.

The causal path analysis was limited because (unlike the HLM), it does not allow us to take into account variance due to students being nested in classes and schools. Not taking into account these sources of variance means that the error variance in the causal path analysis is larger than in the HLM analysis, which reduces the statistical power of the analysis. While this limitation does not affect the internal validity of the results, the lower power of the analysis means that there might be effects in the data that we did not detect. Future research should address this issue by using a larger sample for a causal path analysis.

We also note limitations resulting from the presentation of instructional materials. We conducted our experiment in the context of an intelligent tutoring system, an effective type of educational technology that is widely used in classrooms across the United States. Even though this context represents a realistic educational scenario, further research should test whether our results generalize to out-of-technology contexts. For example, future research should investigate whether our findings generalize to contexts in which students use physical representations or a combination of physical and virtual representations. Further, students received sense-making problems before fluency-building problems. Since this sequence was repeated for each topic of the tutor, we believe that it does not affect the validity of the effects we found in the causal path analysis. However, the effects of fluency-building problems on students' performance on sense-making problems may have been stronger if fluency-building problems had been directly followed by sense-making problems (rather than by individual-representation problems). This limitation may have affected the power of the analysis, but not the validity: we may not have detected all effects, but we can trust the effects that we did detect, and we can trust that the effects we did detect are stronger than the effects we may not have detected.

6. Conclusions

We tested a prediction that resulted from applying KLI to the case of connection making among MGRs; namely, that students will benefit most from support that targets verbally mediated sense-making processes through which students acquire understanding of connections, and support that targets non-verbal, inductive processes through which students acquire perceptual fluency in making connections. Our experiment extends prior research that has only focused on either sense-making support (e.g., Bodemer & Faust, 2006; Seufert, 2003; van der Meij & de Jong, 2011) or fluency-building support (e.g., Kellman & Massey, 203; Kellman et al., 2009), but has not investigated potential interactions between these two types of connection-making support. Our results were more pronounced than expected: the combination of sense-making support and fluency-building support was *necessary* for students to benefit from MGRs, compared to a single GR. The causal path analysis suggests sense-making support provides the foundation for students' benefit from fluency-building support. This finding yields a new testable hypothesis: students will learn best when sense-making support is provided before fluency-building support rather than vice versa.

Given the pervasiveness of MGRs in STEM and the well-documented need for connection-making support, our findings have the potential to apply to many domains. The research presented in this paper is only a first step in this direction, and we hope it will inspire future research on sense making and perceptual fluency in connection making.

References

7. References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. Learning and Instruction, 16(3), 183-198.

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences, 11*(1), 25-61.

Ainsworth, S. E., & van Labeke, N. (2002). *Using a multi-representational design framework to develop and evaluate a dynamic simulation environment*. Paper presented at the International Workshop on Dynamic Visualizations and Learning, Tuebingen, Germany.

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science: A Multidisciplinary Journal, 26*(2), 147-179.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning From Studying Examples to Solving Problems: Effects of Self-Explanation Prompts and Fading Worked-Out Steps. *Journal of Educational Psychology, 95*(4), 774-783.

Barnett, S. M., & Ceci, S. J. (2002). When and Where Do We Apply What We Learn? A Taxonomy for Far Transfer. *Psychological Bulletin, 128*(4), 612-637.

Behr, M. J., Post, T. R., Harel, G., & Lesh, R. (1993). Rational numbers: Toward a semantic analysis - emphasis on the operator construct. In T. P. Carpenter, E. Fennema & T. A. Romberg (Eds.), *Rational Numbers: An Integration of Research*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Berthold, K., Eysink, T. H. S., & Renkl, A. (2008). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science, 27*(4), 345-363. doi: 10.1007/s11251-008-9051-z

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Research, 101*(1), 70-87.

Brinker, L. (1997). Using Structured Representations To Solve Fraction Problems: A Discussion of Seven Students' Strategies (pp. Report: ED409162. 409128p).

Bodemer, D., & Faust, U. (2006). External and mental referencing of multiple representations. *Computers in Human Behavior, 22*(1), 27-42.

Bodemer, D., Ploetzner, R., Bruchmüller, K., & Häcker, S. (2005). Supporting learning with interactive multimedia through active integration of representations. *Instructional Science, 33*(1), 73-95. doi: 10.1007/s11251-004-7685-z

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The Active Integration of Information during Learning with Dynamic and Interactive Visualisations. *Learning and Instruction, 14*(3), 325-341. doi: 10.1016/j.learninstruc.2004.06.006

Bollen, K. A., & Pearl, J. (2013). Eight Myths About Causality and Structural Equation Models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 301-328). The Netherlands: Springer.

Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Lindquist, M. M., & Reys, R. (1980). Results of the second National Assessment of Educational Progress mathematics assessment: Secondary school. *The Mathematics Teacher*, 329-338.

Charalambous, C. Y., & Pitta-Pantazi, D. (2007). Drawing on a Theoretical Model to Study Students' Understandings of Fractions. *Educational Studies in Mathematics, 64*(3), 293-316.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science: A Multidisciplinary Journal, 13*(2), 145-182.

Chickering, D. M. (2002). Optimal Structure Identification with Greedy Search. *Journal of Machine Learning Research 3*, 507-554.

Cook, M., Wiebe, E. N., & Carter, G. (2007). The influence of prior knowledge on viewing and interpreting graphics with macroscopic and molecular representations. *Science Education, 92*(5), 848-867. doi: 10.1002/sce.20262

Corradi, D., Elen, J., & Clareboug, G. (2012). Understanding and Enhancing the Use of Multiple External Representations in Chemistry Education. *Journal of Science Education and Technology, 21*(6), 780-795.

Corwin, R. B., Russell, S. J., & Tierney, C. C. (1990). *Seeing fractions: A unit for the upper elementary grades*: California Dept. of Education.

Cramer, K. (2001). Using models to build an understanding of functions. *Mathematics Teaching in the Middle School, 6*(5), 310-318.

Cramer, K., & Wyberg, T. (2009). Efficacy of different concrete models for teaching the part-whole construct for fractions. *Mathematical Thinking and Learning, 11*, 226-257.

Cramer, K., Wyberg, T., & Leavitt, S. (2008). The role of representations in fraction addition and subtraction. *Mathematics Teaching in the Middle School, 13*(8).

de Jong, T., Ainsworth, S. E., Dobson, M., Van der Meij, J., Levonen, J., Reimann, P., . . . Swaak, J. (1998). Acquiring knowledge in science and mathematics: The use of multiple representations in technology-based learning environments. In M. W. Van Someren, W. Reimers, H. P. A. Boshuizen & T. de Jong (Eds.), *Learning with Multiple Representations*. Bingley, UK: Emerald Group Publishing Limited.

Dreyfus, H., & Dreyfus, S. E. (1986). Five steps from novice to expert *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer* (pp. 16-51). New York: The Free Press.

Eilam, B., & Poyas, Y. (2008). Learning with multiple representations: Extending multimedia learning beyond the lab. *Learning and Instruction, 18*(4), 368-378.

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction, 22*(1), 47-61.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science, 7*, 155-170.

Gibson, E. J. (1969). Principles of perceptual learning and development. New York: Prentice Hall.

Gibson, E. J. (2000). Perceptual learning in development: Some basic concepts. *Ecological Psychology, 12*(4), 295-302.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1-38.

Goldstone, R. (1997). *Perceptual Learning*. San Diego, CA: Academic Press.

Gutwill, J. P., Frederiksen, J. R., & White, B. Y. (1999). Making Their Own Connections: Students' Understanding of Multiple Models in Basic Electricity. *Cognition and Instruction, 17*(3), 249-282.

Hackenberg, A. J. (2007). Units coordination and the construction of improper fractions: A revision of the splitting hypothesis. *The Journal of Mathematical Behavior, 26*(1), 27–47. doi: http://doi.org/10.1016/j.jmathb.2007.03.002

Hattie, J. (2012). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. New York, NY: Routledge.

Hunting, R. P., Davis, G., & Pearn, C. A. (1996). Engaging whole-number knowledge for rational-number learning using a computer-based tool. *Journal for Research in Mathematics Education, 27*(3), 354–379. doi: http://doi.org/10.2307/749369

Johnson, C. I., & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior, 26*(6), 1246-1252.

Jones, L. L., Jordan, K. D., & Stillings, N. A. (2005). Molecular visualization in chemistry education: The role of multidisciplinary collaboration. *Chemistry Education Research and Practice, 6*(3), 136-149. doi: 10.1039/B5RP90005K

Kafai, Y. B., Franke, M. L., Ching, C. C., & Shih, J. C. (1998). Game design as an interactive learning environment for fostering students' and teachers' mathematical inquiry. *International Journal of Computers for Mathematical Learning, 3*(2), 149-184.

Kellman, P. J., & Garrigan, P. B. (2009). Perceptual learning and human expertise. *Physics of Life Reviews, 6*(2), 53-84. doi: 10.1016/j.plrev.2008.12.001

Kellman, P. J., & Massey, C. M. (2013). Perceptual Learning, cognition, and expertise. *The Psychology of Learning and Motivation, 558*, 117-165.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2009). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency *Topics in Cognitive Science* (Vol. 2, pp. 285-305).

Kieren, T. E. (1993). Rational and fractional numbers: From quotient fields to recursive understanding. In T. P. Carpenter, E. Fennema & T. A. Romberg (Eds.), *Rational numbers: An integration of research*. Hillsdale, NJ: Erlbaum.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional Complexity and the Science to Constrain It. *Science Education, 342*(6161), 935-937. doi: doi:10.1126/science.1238056 doi:doi:10.1126/science.1238056

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science, 36*(5), 757–798.

Kozma, R., & Russell, J. (2005). Students becoming chemists: Developing representationl competence. In J. Gilbert (Ed.), *Visualization in science education* (pp. 121-145). Dordrecht, Netherlands: Springer.

Lamon, S. J. (Ed.). (1999). *Teaching Fractions and Ratios for Understanding*. Mahwah, NJ: Lawrence Erlbaum Associates.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science: A Multidisciplinary Journal, 11*(1), 65-100.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., & Busick, M. D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. *Institute of Education Sciences, National Center for Special Education Research*, Downloaded from http://ies.ed.gov/ncser/pubs/20133000/ on 20133012/20133018/20132012.

Lobato, J., Ellis, A., & Zbiek, R. M. (2010). Developing essential understanding of ratios, proportions, and proportional reasoning for teaching mathematics: Grades 6-8. Reston, VA: National Council of Teachers of Mathematics.

Mack, N. (1995). Confounding whole-number and fraction concepts when building on informal knowledge. *Journal for Research in Mathematics Education, 26*(5), 422-441.

Martin, W. G., Strutchens, M. E., & Elliott, P. C. (2007). *The Learning of Mathematics, 69th NCTM Yearbook*. Reston, VA: National Council of Teachers of Mathematics.

Martinie, S. L., & Bay-Williams, J. M. (2003). Investigating Students' Conceptual Understanding of Decimal Fractions Using Multiple Representations. *Mathematics Teaching in the Middle School, 8*(5), 244-247.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 31-48). New York, NY: Cambridge University Press.

Moss, J. (2005). Pipes, Tubes, and Beakers: New approaches to teaching the rational-number system. In J. Brantsford & S. Donovan (Eds.), *How people learn: A targeted report for teachers* (pp. 309-349): National Academy Press.

Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: A new model and an experimental curriculum. *Journal for Research in Mathematics Education, 30*(2), 122-147.

National Mathematics Advisory Panel (2008). Foundations for Success: Report of the National Mathematics Advisory Board Panel: U.S. Government Printing Office.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2006). Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence. Reston, VA.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA, US: Harvard University Press.

Ni, Y., & Zhou, Y.-D. (2005). Teaching and Learning Fraction and Rational Numbers: The Origins and Implications of Whole Number Bias. *Educational Psychologist, 40*(1), 27-52.

Ohlsson, S. (1988). Mathematical meaning and applicational meaning in the semantics of fractions and related concepts. In J. Hiebert & M. Behr (Eds.), *Research agenda for mathematics education: Number concepts and operations in the middle grades* (pp. 53–92). Reston, VA: National Council of Teachers of Mathematics.

Özgün-Koca, S. A. (2008). Ninth grade students studying the movement of fish to learn about linear relationships: The use of video-based analysis software in mathematics classrooms. *The Mathematics Educator, 18*(1), 15-25.

Pape, S. J., & Tchoshanov, M. A. (2001). The role of representation (s) in developing mathematical understanding. *Theory into Practice, 40*(2), 118-127. doi: 10.1207/s15430421tip4002_6

Patel, Y., & Dexter, S. (2014). Using Multiple Representations to Build Conceptual Understanding in Science and Mathematics. In M. Searson & M. Ochoa (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2014* (Vol. 2014, pp. 1304-1309). Chesapeake, VA: AACE.

Pearl, J. (2000). *Causality: Models, Reasoning, and Interference*: Cambridge University Press.

Pitta-Pantazi, D., Gray, E. M., & Christou, C. (2004). Elementary School Students' Mental Representation of Fractions. *International Group for the Psychology of Mathematics Education*.

Rau, M. A., Rummel, N., Aleven, V., Pacilio, L., & Tunc-Pekkan, Z. (2012a). How to schedule multiple graphical representations? A classroom experiment with an intelligent tutoring system for fractions. In J. v. Aalst, K. Thompson, M. J. Jacobson & P. Reimann (Eds.), *The future of learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012).* (Vol. 1, pp. 64-71). Sydney, Australia: ISLS.

Rau, M. A., Aleven, V., Rummel, N., & Rohrbach, S. (2012). Sense making alone doesn't do it: Fluency matters too! ITS support for robust learning with multiple representations. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems, ITS 2012* (Vol. 7315, pp. 174-184). Berlin: Springer.

Rau, M. A., Aleven, V., Rummel, N., & Rohrbach, S. (2013). Why interactive learning environments can have it all: Resolving design conflicts between conflicting goals. In W. E. Mackay, S. Brewster, & S. Bødker (Eds.), *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2013)* (pp. 109-118). New York: ACM.

Rau, M. A., Aleven, V., Rummel, N., & Pardos, Z. (2014). How Should Intelligent Tutoring Systems Sequence Multiple Graphical Representations of Fractions? A Multi-Methods Study. *International Journal of Artificial Intelligence in Education, 24*(2), 125-161. doi: 10.1007/s40593-013-0011-7

Rau, M. A., Aleven, V., & Rummel, N. (2015). Successful learning with multiple graphical representations and self-explanation prompts. *Journal of Educational Psychology,107*(1), 30-46. doi: 10.1037/a0037211

Raudenbush, S. W., & Bryk, A. S. (Eds.). (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage Publications.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23*, 90-108.

Richman, H. B., Gobet, F., Staszewski, J. J., & Simon, H. A. (1996). Perceptual and Memory Processes in the Acqquisition of Expert Performance: The EPAM Model. In K. A. Ericsson (Ed.), *The road to excellence? The acqquisition of expert performance in the arts and sciences, sports and games* (pp. 167-187). Mahwah, NJ: Erlbaum Associatees.

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 49-69). New York, NY: Cambridge University Press.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*(2), 141-156.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122*(2), 166-183.

Schooler, J. W., Fiore, S., & Brandimonte, M. A. (1997). At a loss From words: Verbal overshadowing of perceptual memories. *Psychology of Learning and Motivation: Advances in Research and Theory, 37*, 291–340.

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of learning: Research and perspectives* (pp. 1-51). Greenwich, CT: Information Age Publishing.

Seufert, T. (2003). Supporting Coherence Formation in Learning from Multiple Representations. *Learning and Instruction, 13*(2), 227-237.

Seufert, T., & Brünken, R. (2006). Cognitive load and the format of instructional aids for coherence formation. *Applied Cognitive Psychology, 20*, 321-331. doi: 10.1002/acp.124810.1002/acp.124810.1002/acp.124810.1002/acp.124810.1002/acp.1248

Shanks, D. (2005). Implicit learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of Cognition* (pp. 202-220). London: Sage.

Siegler, R. S., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., . . . Wray, J. (2010). *Developing effective fractions instruction: A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2013). Fractions: the new frontier for theories of numerical development. *Trends in Cognitive Sciences, 17*(1), 13–19. doi: 10.1016/j.tics.2012.11.004

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62*(4), 273-296

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction, and Search* (2 ed.): MIT Press.

Steffe, L. P. (2002). A new hypothesis concerning children's fractional knowledge. *The journal Of Mathamtical Behavior, 20*(3), 267-307. doi: http://doi.org/10.1016/S0732-3123(02)00075-5

Stern, E., Aprea, C., & Ebner, H. G. (2003). Improving cross-content transfer in text processing by means of active graphical representation. *Learning and Instruction, 13*(2), 191-203.

Taber, S. B. (2001). Making Connections among Different Representations: The Case of Multiplication of Fractions.

Thompson, D. R., & Saldanha, L. A. (2003). Fractions and multiplicative reasoning. In J. Kilpatrick, G. Martin & D. Schrifter (Eds.), *Research companion to the principles and standards for school mathematics*. Reston, VA: National Council of Teachers in Mathematics.

Van der Meij, J., & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction, 16*(3), 199-212. doi: 10.1016/j.learninstruc.2006.03.007

Van der Meij, J., & de Jong, T. (2011). The effects of directive self-explanation prompts to support active processing of multiple representations in a simulation-based learning environment. *Journal of Computer Assisted Learning, 27*(5), 411-423.

van Merrienboër, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Etr&D-Educational Technology Research and Development, 50*(2), 39-64.

Van Merrienboër, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(1), 147-177.

Van Someren, M. W., Boshuizen, H. P. A., & de Jong, T. (1998). Multiple representations in human reasoning. In M. W. Van Someren, H. P. A. Boshuizen & T. de Jong (Eds.), *Learning with multiple representations*. Oxford: Pergamon.

Vreman-de Olde, C., & De Jong, T. (2007). Scaffolding learners in designing investigation assignments for a computer simulation. *Journal of Computer Assisted Learning, 22*, 63-73. doi: 10.1111/j.1365-2729.2006.00160.x

Wise, J. A., Kubose, T., Chang, N., Russell, A., & Kellman, P. J. (2000). Perceptual learning modules in mathematics and science instruction. In P. Hoffman & D. Lemke (Eds.), *Teaching and Learning in a Network World* (pp. 169-176): IOS Press.

Wu, H. K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education, 88*(3), 465-492. doi: 10.1002/sce.10126

Wylie, R., Koedinger, K. R., & Mitamura, T. (2009). Is self-explanation always better? The effects of adding self-explanation prompts to an English grammar tutor. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1300–1305). Amsterdam, The Netherlands.

Yang, D.-C., & Reys, R. E. (2001). One fraction problem, many solution paths. *Mathematics Teaching in the Middle School, 7*(3), 164-166.

Tables

Table 1. Overview of experimental conditions.

		Sense-making support			**Control**
		No	*Linked representations*	*Analogous examples*	
Fluency-building support	*No*	Multiple-graphical-representations (MGR)	Sense-making with linked GRs (SL)	Sense-making with analogous examples (SE)	
	Yes	Fluency-building (F)	Sense-making with linked GRs plus fluency-building (SL-F)	Sense-making with analogous examples plus fluency-building (SE-F)	
Control					Single-graphical-representation (SGR)

Table 2. Problem sequence per condition: for each topic, problems 1-4 (P1-P4) are individual-representation problems (I); problems 5-8 are connection-making problems: sense-making problems with analogous examples (SE, underlined) or perceptual fluency-building problems (F, italicized). Bold-underlined problems and bold-italicized problems are used in the causal path analysis.

Condition	Topic	P1	P2	P3	P4	P5	P6	P7	P8
SE	1	I	I	I	I	**SE**	**SE**	SE	SE
	2	I	I	I	I	**SE**	**SE**	SE	SE
	…					…			
F	1	I	I	I	I	*F*	*F*	***F***	***F***
	2	I	I	I	I	*F*	*F*	***F***	***F***
	…					…			
SE-F	1	I	I	I	I	**SE**	**SE**	***F***	***F***
	2	I	I	I	I	**SE**	**SE**	***F***	***F***
	…					…			

Table 3. Overview of variables included in the HLM.

Variable	Explanation
Y_{ijkl}	Student$_i$'s score on the dependent measures at test time t_i (i.e., immediate or delayed posttest)
ε_{jkl}	Intercept for student$_j$'s score
β_1	Effect of test time t_i
β_2	Effect of sense-making problems with prompts for analogical comparisons GRs se_j
β_3	Effect of sense-making problems with linked GRs sl_j
β_4	Effect of perceptual fluency-building problems f_j
β_5	Effect of student$_j$'s performance on the pretest p_j
β_6	Aptitude-treatment interaction between sense-making problems with analogical comparisons se_j and student$_j$'s performance on pretest p_j
β_7	Aptitude-treatment interaction between sense-making problems with linked GRs sl_j and student$_j$'s performance on pretest p_j
β_8	Aptitude-treatment interaction between perceptual fluency-building problems f_j and student$_j$'s performance on pretest p_j
δ_{kl}	Random intercept for class$_k$
γ_l	Random intercept for school$_l$
μ	Overall average

Table 4. Means (and standard deviations) for conceptual and procedural knowledge at pre-test, immediate posttest, delayed posttest. Min. score is 0, max. score is 1.

Measure	Condition	Pretest	Immediate posttest	Delayed posttest
Conceptual knowledge	Multiple-graphical-representations (MGR)	.33 (.20)	.45 (.23)	.48 (.26)
	Sense-making with linked GRs (SL)	.38 (.20)	.49 (.23)	.51 (.26)
	Sense-making with analogous examples (SE)	.36 (.22)	.43 (.20)	.49 (.26)
	Fluency-building (F)	.31 (.21)	.37 (.22)	.44 (.24)
	Sense-making with linked representations plus fluency-building problems (SL-F)	.36 (.20)	.43 (.24)	.49 (.25)
	Sense-making with analogous examples plus fluency-building problems (SE-F)	.39 (.21)	.52 (.24)	.58 (.26)
	Single-graphical-representation (SGR)	.37 (.20)	.43 (.25)	.48 (.20)
Procedural knowledge	Multiple-graphical-representations (MGR)	.25 (.25)	.30 (.28)	.30 (.26)
	Sense-making with linked representations (SL)	.21 (.18)	.26 (.24)	.26 (.24)

	Sense-making with analogous examples (SE)	.26 (.21)	.29 (.24)	.31 (.27)
	Fluency-building condition (F)	.19 (.17)	.23 (.20)	.25 (.22)
	Sense-making with linked representations plus fluency-building problems (SL-F)	.20 (.18)	.25 (.21)	.26 (.21)
	Sense-making with analogous examples plus fluency-building problems (SE-F)	.26 (.20)	.32 (.26)	.33 (.26)
	Single-graphical-representation (SGR)	.21 (.20)	.25 (.22)	.27 (.23)

Table 5. Means (and standard deviations) of total time spent on tutor problems by condition.

Condition	Time on tutor in minutes
Multiple-graphical-representations (MGR)	232.04 (62.88)
Sense-making with linked GRs (SL)	206.27 (60.3)
Sense-making with analogous examples (SE)	213.7 (58.32)
Fluency-building (F)	199.25 (54.97)
Sense-making with linked representations plus fluency-building problems (SL-F)	215.83 (58.43)
Sense-making with analogous examples plus fluency-building problems (SE-F)	203.51 (53.61)
Single-graphical-representation (SGR)	189.47 (41.54)

Table 7. Error types on fluency-building problems and number of occurrences per condition (summed up for all students across fluency-building problems P7 and P8). Italicized error types were selected for further analysis.

Error type	**Knowledge component**	**# in F**	**# in SE-F**
nameCircleMixed-Error	Finding circle representations that show the same fraction as a number line or a rectangle	355	126
equivalenceError	Finding equivalent fraction representations	2899	2157
improperMixed-Error	Finding representations of improper fractions	1380	1608
additionMixedError	Finding representations that show the addend of a given sum equation depicted by representations	207	176
compareMixed-Error	Finding representations that show a fraction smaller or larger than the given one	436	307

diffMixedError	Finding representations that show the difference of two fractions	282	238
nameNLMixed-Error	Finding number line representations that show the same fraction as a circle or a rectangle	949	599
nameRectMixed-Error	Finding rectangle representations that show the same fraction as a number line or a circle	385	133
subtractionMixed-Error	Finding representations that show the subtrahend of a given difference equation depicted by representations	214	240
sumMixedError	Finding representations that show the sum of two fractions	256	205
unitMixedError	Finding the unit of a given fraction	1050	1138

Table 8. Error types on sense-making problems and number of occurrences per condition (summed up for all students across sense-making problems P5 and P6). Italicized error types were selected for further analysis.

Error type	Knowledge component	# in SE	# in SE-F
place1Error	Locating 1 on the number line given a dot on the number line and the fraction it shows	150	222
selfExplanationError	Incorrect response to self-explanation prompt	1320	1629
comparisonError	Comparing two fractions	92	82
denomError	Entering the denominator of a fraction	972	837
equivalence-CompareError	Judging whether two fractions are equivalent	19	18
multiplyError	Entering a number by which to multiply numerator or denominator to expand a given fraction	30	29

nlPartitionError	Partitioning the number line to show an equivalent fraction	1913	2115
numberSections-UnitError	Finding the denominator of a fraction by indicating how many sections the unit was divided into	41	44
numError	Entering the numerator of a fraction	1559	1390
placeDotError	Placing a dot on the number line to show a fraction	198	253
sectionsBetween-0-1	Indicating that the denominator in a number line is shown by the sections between 0 and 1	61	44
unitError	Selecting the unit for a fraction given the symbolic fraction and a graphical representation	123	115

Figure

Figures

Figure 1. Graphical representations of fractions: circle, rectangle, and number line.

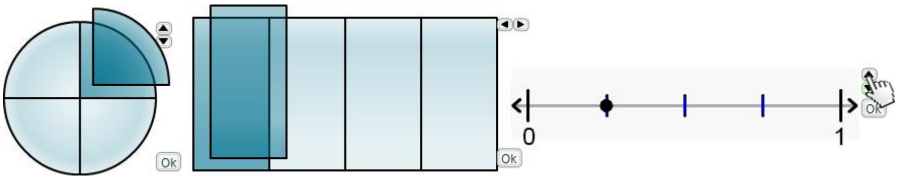

Figure 2. Theory of change of how working on connection-making problems (sense-making problems, fluency-building problems) foster learning processes (verbally mediated sense-making processes, non-verbal inductive and refinement processes) and representational competences (understanding of connections and perceptual fluency in making connections) that enhance students' learning of robust domain knowledge (robust fractions knowledge). For each mechanism, the figure indicates which section in the paper describes prior research regarding this particular mechanism.

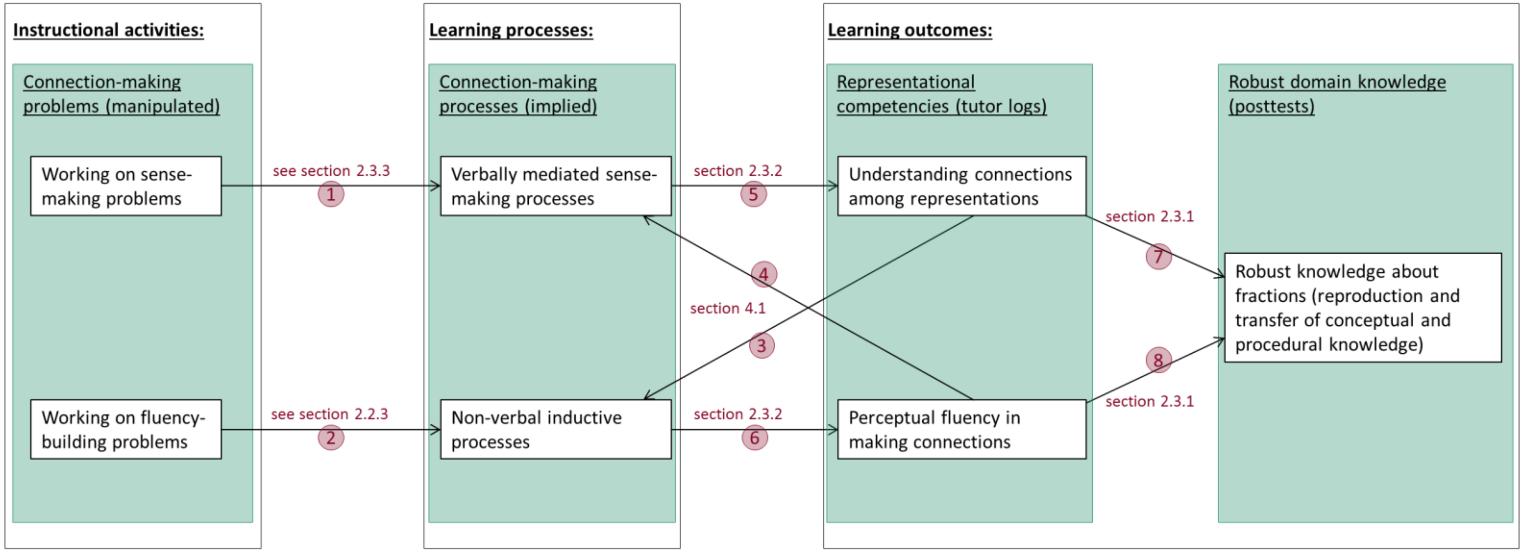
Instructional activities:
Connection-making problems (manipulated)
Working on sense-making problems
Working on fluency-building problems
Learning processes:
Connection-making processes (implied)
Verbally mediated sense-making processes
Non-verbal inductive processes
Learning outcomes:
Representational competencies (tutor logs)
Understanding connections among representations
Perceptual fluency in making connections
Robust domain knowledge (posttests)
Robust knowledge about fractions (reproduction and transfer of conceptual and procedural knowledge)
see section 2.3.3
1
see section 2.2.3
2
section 2.3.2
5
section 2.3.2
6
section 4.1
3
4
section 2.3.1
7
section 2.3.1
8

Figure 3. Example of a tutor problem with only the number-line representation.

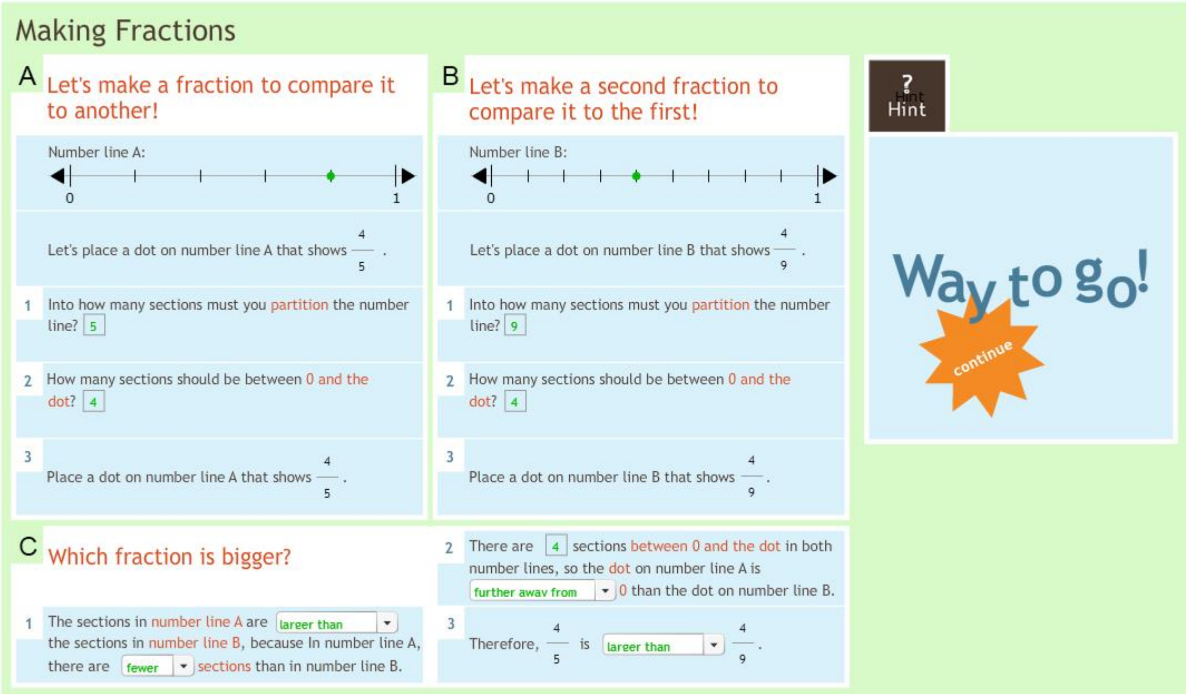

Figure 4. Example of a sense-making problem with analogous examples. The self-explanation prompts in part C (highlighted in pink) were identical to sense-making problems with linked representations.

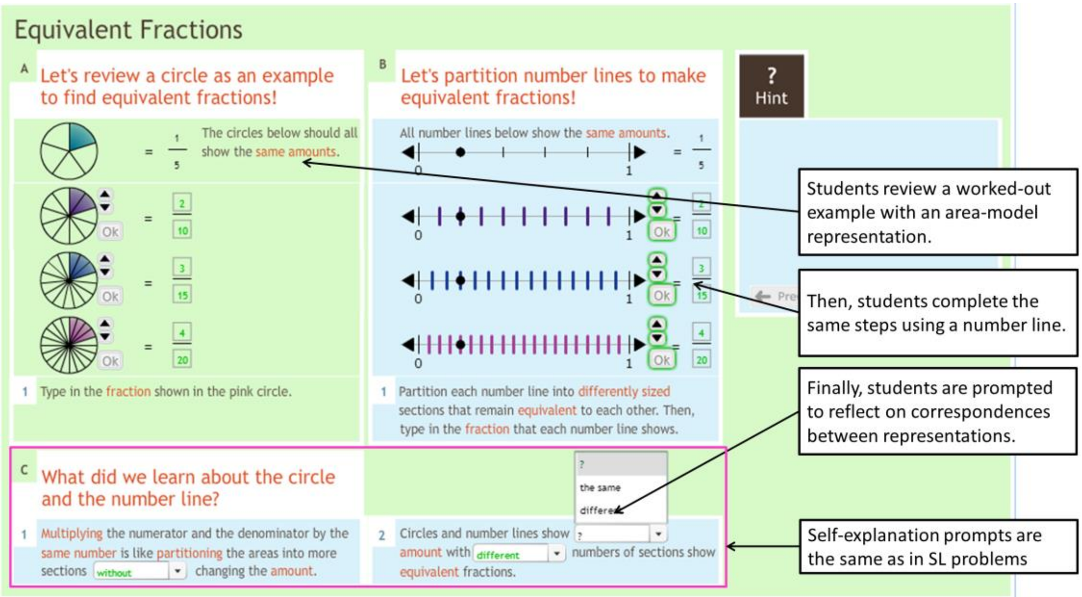

Figure 5. Example of sense-making problem with linked representations. The self-explanation prompts in part B (highlighted in pink) were identical to sense-making problems with analogous examples.

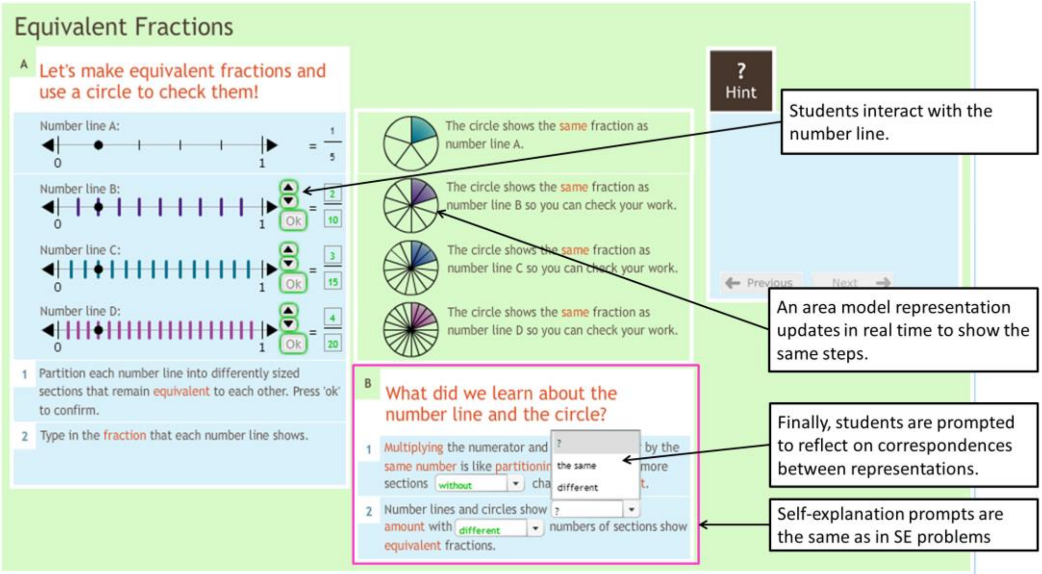

Figure 6. Example of a fluency-building problem.

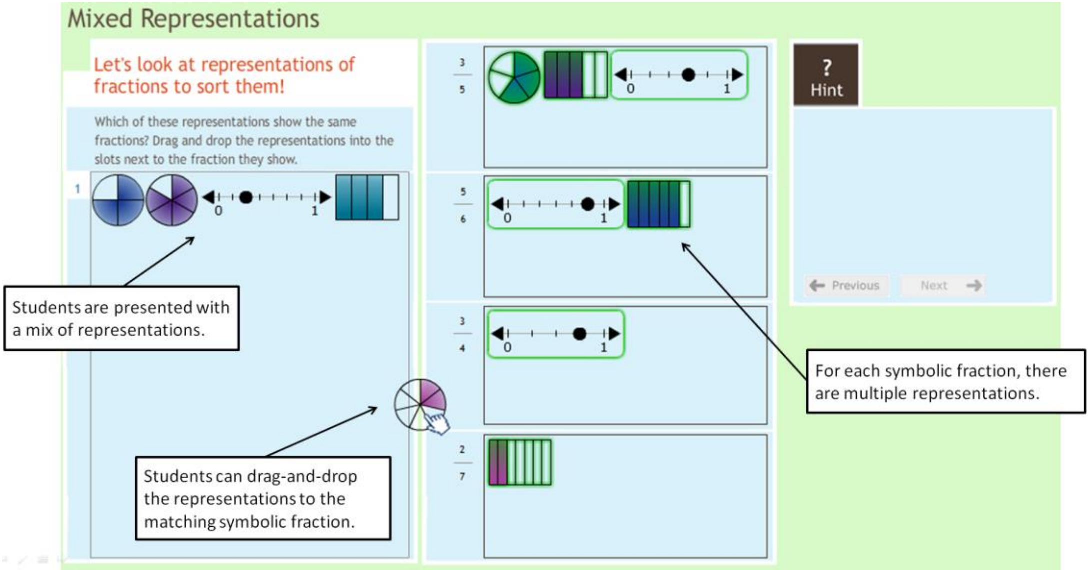

Figure 7. Saturated models for the fluency hypothesis (left) and the sense-making hypothesis (right).

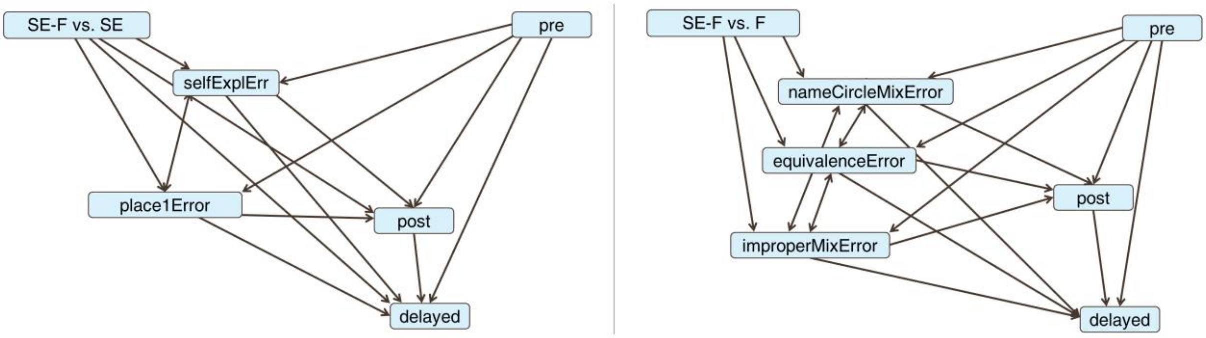

Figure 8. Fluency-hypothesis model with unstandardized parameter estimates. Paths that describe a negative effect of fluency-building support on posttest performance (immediate and final) are highlighted in red, paths that describe a positive effect are highlighted in green.

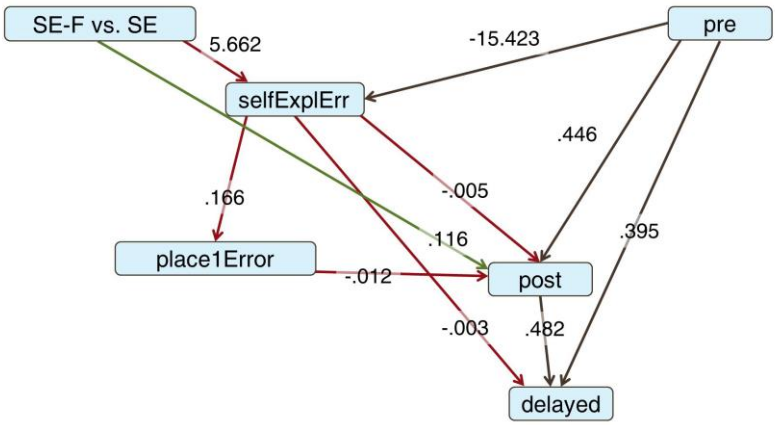

Figure 9. Sense-making hypothesis model with unstandardized parameters.