How can we promote understanding and fluency in learning from multiple representations? Intelligent Tutoring System support for connection making
Martina Rau

Committee members: Vincent Aleven, Nikol Rummel, Ken Koedinger, Richard Scheines, & Brian Junker

Abstract: Multiple graphical representations (MGRs) can improve students’ learning. But students’ benefit from them depends on their ability to make connections between graphical representations (GRs). Based on several studies with an intelligent tutoring system (ITS), I have developed a theoretical model for learning with MGRs. To benefit from MGRs, students need to engage in both sense-making processes to understand connections between GRs and they need to become fluent in making these connections. I propose to investigate how these crucial learning processes interact: Does understanding facilitate fluency-building processes or does fluency enhance sense-making processes? And, consequently, which process should ITSs support first? According to the fluency-first hypothesis, providing students with fluency-building support before sense-making support frees cognitive resources that students need to benefit from sense-making support. By contrast, the understanding-first hypothesis predicts that sense-making support equips students with the necessary knowledge about structural correspondences between GRs, allowing them to attend to relevant aspects of the GRs while solving fluency-building problems. Both hypotheses make predictions about students’ efficiency in directing their attention to relevant features of learning materials, about their ability to make conceptual connections, and about their performance during the learning phase. I propose to analyze these mediation effects on students’ learning using eye-tracking, cued retrospective reports, and tutor log data. I will conduct the experiment in the context of an ITS that I have developed and iteratively improved as part of my PhD research.

The proposed experiment will contribute to the multiple representations literature as it investigates how different types of instructional support interact, while investigating mediation effects by integrating eye-tracking data, cued retrospective report data, and tutor log data. It will extend the cognitive science literature by investigating how different kinds learning processes interact: sense-making processes and fluency-building processes. It will lead to instructional principles for combining support for understanding of and fluency in connection making, thereby contributing to the ITS literature. Finally, the findings of the proposed research have the potential to help develop cognitive models that can be used to trace students’ understanding of connections and their fluency in making connections for the purpose of adaptive tutoring.
# 1 Introduction

Instructional materials almost universally employ multiple representations: flow diagrams are used in programming, schemas and tree diagrams in biology, charts and diagrams in mathematics – to mention only a few examples. External representations are considered to be useful instructional tools because they can clarify crucial aspects of the learning content, often through the use of perceptually intuitive characteristics (Ainsworth, 2006). Since different representations often emphasize complementary aspects of the learning content (Ainsworth, 2006), instructional materials tend to use not only one representation, but multiple. Indeed, in the educational psychology literature, there is extensive evidence that learning with multiple representations can enhance conceptual learning of the domain (Ainsworth, 2006; Ainsworth, Bibby, & Wood, 1998; Berthold, Eysink, & Renkl, 2008; de Jong et al., 1998; Lohner, van Joolingen, & Savelsbergh, 2003; Schnitz & Bannert, 2003; Van der Meij, 2007; Van der Meij & de Jong, 2006).

By and large, the educational psychology literature on learning with multiple representations has taken a dual-representation approach: research has mostly focused on learning with textual descriptions and one additional graphical representation (GR) (Bodemer & Faust, 2006; Bodemer, Plötzner, Feuerlein, & Spada, 2004; Butcher, 2006; Schnitz & Bannert, 2003). Many instructional materials found in real educational settings, however, are more complex: these materials contain multiple graphical representations (MGRs) in addition to text and symbolic representations. Fractions instruction is one of the many domains in which MGRs, such as circles, rectangles, and number lines are used extensively (NMAP, 2008). Different GRs emphasize different conceptual aspects of fractions (Charalambous & Pitta-Pantazi, 2007). For instance, area models (i.e., circles and rectangles) depict fractions as equally sized parts of a whole, where the whole is usually inherent to the shape (e.g., the whole circle in a circle representation). Students interpret area models by relating the number of colored sections to the number of total sections in the unit. They can compare the relative size of fractions represented in area models by comparing the relative colored area to the whole area of the shape. Area models are often used in the context of sharing activities, thus building on students’ intuitive knowledge about fractions. Linear models (e.g., number lines) depict fractions in the context of measurement and demonstrate that fractions can lie between any two whole numbers. Linear models do not have an inherent unit, but the unit is defined by convention: the length between 0 and 1 in is the unit, as opposed to the length of the entire number line (e.g., in a number line from 0 to 3). To compare the relative size of fractions using linear models, students have to judge the length of a linear segment relative to the defined unit of the representation. The goal in using MGRs is to help students understand the complex topic of fractions by highlighting these complementary conceptual aspects. This practice makes fractions a suitable domain to research learning with MGRs: as in many other STEM domains, various GRs, each with a different conceptual focus, are used in conjunction with textual descriptions and symbolic representations to enhance learning of a rather abstract concept.

In spite of the well-documented promise of learning with multiple representations, research has not always succeeded in demonstrating their advantage on students’ learning
(Ainsworth, 2006; Bodemer, Ploetzner, Bruchmüller, & Häcker, 2005; de Jong, et al., 1998; van der Meij & de Jong, 2006). Unfortunately, we do not yet fully understand the mechanisms by which multiple representations enhance learning. Without such knowledge, however, we cannot develop appropriate instructional design principles for the development of effective instructional materials that promote robust learning of a domain: learning of flexible knowledge that students can transfer to novel tasks and that lasts over time (Koedinger, Corbett, & Perfetti, in press).

The proposed research addresses open questions about the processes involved in successful learning from multiple representations and how to provide instructional support for these processes. Further, my research extends prior research on learning with dual representations to learning with multiple graphical representations, a common scenario in real-world instructional materials. To achieve these goals, I have conducted several experiments on learning with MGRs of fractions which I propose to be part of my thesis statement (Rau, Aleven, & Rummel, 2010; Rau, Aleven, & Rummel, in press; Rau, Aleven, Rummel, & Rohrbach, 2012; Rau, Rummel, Aleven, Tunc-Pekkan, & Pacilio, 2012). The goal of each experiment was to (1) investigate whether a hypothesized learning process is involved in successful learning with MGRs, (2) investigate how best to design instructional support for this process, and (3) develop and further refine an Intelligent Tutoring System (ITS) for fractions learning that incorporates these instructional design principles. Based on this research, and based on existing theoretical models in the educational psychology literature, I have developed a theoretical model for learning with MGRs. I propose to conduct a final experiment to investigate the interactions between instructional support for two crucial learning processes involved in making connections between MGRs: sense-making processes and fluency-building processes. Sense-making processes are defined as “explicit, verbally-mediated learning in which students attempt to understand or reason” (Koedinger et al., in press, p. 30). Sense-making processes lead to the understanding of KCs (structural components of the domain knowledge, such as numerator, denominator, and unit in fractions) by relating them to their underlying principles. Fluency-building processes are “non-verbal learning processes involved in strengthening memory and compiling knowledge, producing more automatic and composed (‘chunked’) knowledge” (Koedinger et al., in press, p. 29). Fluency with representations refers to the ability to “[…extract] information more quickly and automatically with practice” (Kellman and Garigan, 2009; p. 3), and to “[…see] at a glance what is relevant […] and to ignore what is not” (Kellman et al., 2008, p. 358).

Before describing the planned experiment in detail, I will discuss relevant theoretical background, my own theoretical model for learning with MGRs, and my own research that provides empirical evidence for this model.

1.1 Models for Learning with Multiple External Representations
When discussing theoretical models for learning with multiple representations, one needs to distinguish internal and external representations. External representations are “the knowledge and structure in the environment, as physical symbols, objects, or dimensions […] embedded in physical configurations” (Zhang, 1997; p. 180). Internal representations, on the other hand, are
knowledge structures in memory, such as schemas or production rules (Zhang, 1997). As learners understand external representations, they form internal representations, which (ideally) they will then integrate into a mental model (Zhang, 1997; Zhang & Norman, 1994). Since different representations may have complementary strengths (Ainsworth, 2006), the effectiveness of multiple external representations lies in their potential to help students form appropriate mental models of the domain.

In line with the Dual Channel theory (Paivio, 1986), Mayer and Moreno’s (2003) Cognitive Theory of Multimedia Learning assumes that verbal and pictorial information are processed in different information channels. Since the capacity of each information channel is limited, but the capacity of both channels together is additive (Chandler & Sweller, 1991), learning with both text and GR makes better use of the learner’s mental capacity. Furthermore, active integration of the textual and picture-based internal representations into a coherent mental model requires deeper conceptual processing of the content, which leads to better learning.

Building on the Cognitive Theory of Multimedia Learning, Schnotz and Bannert’s (2003) theoretical account for learning from dual representations (Fig. 1) proposes that text and GR lead to different types of internal representations. Text is processed semantically through an analysis of its symbolic structure, leading to a propositional internal representation. The GR, on the other hand, is processed perceptually, leading to a pictorial internal representation. During mental model formation, learners integrate both internal representations via structure mapping. The integration process of information from different sign systems requires deep conceptual processing, which explains better learning from text and GR than from text alone.

![Schnotz & Bannert's (2003) theoretical model of text and graphic comprehension.](image)

While the Cognitive Theory of Multimedia Learning (Mayer & Moreno, 2003) and the theoretical model by Schnotz and Bannert (2003) can explain why dual representations (i.e., text and one GR) lead to better learning, they do not predict an advantage of MGRs compared to a single GR. Since all GRs are processed in the visual information channel, MGRs should not increase a learner’s cognitive capacity. In contrast, MGRs should increase the chances of
cognitive overload in the visual channel, which might hamper learning. Furthermore, since MGRs share the same sign system and organization principles, conceptual processing resulting from an integration of information from different sign systems does not come into play.

However, the Design-Functions-Tasks framework (DeFT; Ainsworth, 2006), explains why MGRs might yet enhance learning. Through the function of computational offloading, multiple representations can reduce the amount of cognitive effort required to process one representation by providing another. For instance, number lines and circles may represent the same fraction, but the number line is more difficult to interpret than the circle, because it uses more abstract features (e.g., labels of 0 and 1 at the ends of the number line to denote the unit of the fraction, rather than having an inherit unit, the whole circle). Providing the circle along with the number line may thus help a learner understand the number line. Re-representation describes the function of different representations to emphasize different aspects of the concept they represent, even though the representations might have the same abstract structure. For instance, a circle may emphasize that a fraction (which is usually smaller than 1) is a part of a whole, whereas a number line emphasizes that a fraction can fall between any two whole numbers (not only between 0 and 1), although they share the same structure: both number line and circle depict the KCs of numerator and denominator. Finally, the function of graphical constraining describes that one representation may limit the interpretation of another. For example, a circle provided along with a number line may help a student interpret the number line correctly. Consider the case that a circle shows 1/2, and a number line shows a segment from 0 to 2, with a dot at 1/2. A student might apply the part-whole approach to the number line and interpret the dot as showing 1/4 (i.e., taking the entire number line as the unit, rather than just the segment between 0 and 1). Knowing that the circle and the number line are both supposed to show the same fraction (i.e., 1/2) may help the student overcome the misconception that the entire segment shown by a number line (as opposed to the segment of just 0 to 1) denotes the unit. DeFT further describes cognitive tasks that learners need to accomplish when learning with multiple external representations. Learners need to understand the form of each representation (i.e., they have to learn how a representation depicts information). They have to learn how to use the representation within the domain and how to construct the representation. Finally, they need to know how to select an appropriate representation for a given task, for which the ability to make connections between representations and to compare them to one another is an important prerequisite.

However, DeFT does not specify the learning processes necessary to benefit from MGRs, or how to provide instructional support for them. Building on the prior work described in this section, I propose a new theoretical model that specifies the learning processes students need to engage in so as to benefit from MGRs.

2 A Model for Learning with Multiple Graphical Representations
In almost any domain, learning involves sense-making processes. Sense-making processes are learning processes that lead to principled understanding of connections between MGRs based on their knowledge components (KCs; e.g., numerator and denominator). In order to effectively use
MGRs to learn about a domain, students need to engage in *sense-making processes* to develop *understanding of each individual GR* and *understanding of the connections between MGRs*.

*Sense-making processes* in *understanding individual GRs* involve the ability to relate the KCs involved in each GR (e.g., the number of colored sections) to the abstract concept they represent (e.g., the numerator). This notion of understanding individual GRs includes understanding of its format (Ainsworth, 2006), understanding of the operators a GR uses (Ainsworth, 2006), understanding of the relation between the GR and the domain (Ainsworth, 2006), and the ability to use the GR to solve a task in the domain (de Jong, et al., 1998; Nistal et al., 2009). Understanding GRs has been shown to be a difficult task in many domains (Baker, Corbett, & Koedinger, 2002; Friel, Curcio, & Bright, 2001; Kaput, 1989; Preece, 1993), and is generally recognized as an important educational goal in fractions (Siegl et al., 2010; NCTM, 2010), geometry and algebra (NMAP, 2008).

The bottom part of Fig. 2 illustrates these sense-making processes in developing *understanding of individual GRs* using three GRs of fractions as an example. Here, students are presented with external GRs (e.g., a circle, a rectangle, and a number line). To develop understanding of each individual GR, students need to relate each GR to the domain-specific KCs it depicts. To this end, students need to learn what component of each GR corresponds to the KCs numerator, denominator, and unit of the fraction: in the circle and rectangle, the numerator corresponds to the number of colored sections, the denominator to the total number of sections, and the unit to the inherent shape of the representation. For the number line, the numerator is the number of sections between 0 and the dot, the denominator the number of sections between 0 and 1, and the unit is always the length between 0 and 1. During this sense-making process, students form an internal representation of the KCs depicted in each GR.

*Sense-making processes* in *making connections* between MGRs lead to *understanding* how different GRs relate to one another. Making sense of connections requires students to relate corresponding KCs of different GRs. For example, a student might reason that a circle with one colored section shows the numerator of 1, and that a number line with one section between 0 and the dot shows a numerator of 1; that a circle with two total sections shows a denominator of 2, that the number line with two sections between 0 and 1 shows a denominator of 2; and that, since both GRs show the same numerator and the same denominator, they both show 1/2. The student reasoned on the basis of the connections between the GRs at the level of the KCs numerator and denominator. The top part of Fig. 2 illustrates the *sense-making processes* for connection making between MGRs of fractions just described. The ability to make connections between multiple representations has been demonstrated to be key to students’ benefit from them (Ainsworth, 2006; Bodemer & Faust, 2006; Bodemer, et al., 2005; Bodemer, et al., 2004; Brünken, Seufert, & Zander, 2005; Butcher & Aleven, 2008; Gutwill, Frederiksen, & White, 1999; Plötzner, Bodemer, & Feuerlein, 2001; Taber, 2001; van der Meij & de Jong, 2006).

Learning does not only rely on understanding: knowledge is only useful if it is readily accessible whenever needed. A learner who has readily accessible knowledge is said to have *fluency* in that knowledge. In addition to understanding fractions representations, being fluent
Figure 2. Theoretical model for sense-making processes in developing knowledge about individual graphical representations and knowledge about connections between multiple graphical representations.

Figure 3. Theoretical model for fluency-building processes in developing knowledge about individual graphical representations and knowledge about connections between multiple graphical representations.
with fractions, in representing fractions and in relating different representations of fractions has been recognized as an important foundation of Algebra learning (NMAP, 2008). Fluency-building processes result from experience with the perceptual properties of GRs and lead to readily accessible perceptual knowledge about individual GRs and about the connections between MGRs.

Accordingly, fluency with individual GRs describes the ability to quickly and effortlessly identify the information it shows and to use it to solve a task in the domain. The bottom part of Fig. 3 shows these fluency-building processes in developing understanding of individual GRs of fractions. Fluency with individual GRs means that students associate the abstract fraction shown by each GR without reasoning about the KCs of numerator and denominator separately. Instead, they treat each GR as one perceptual chunk that stands for a given fraction.

A student who is fluent in making connections between MGRs can quickly and effortlessly relate different GRs by judging “at a glance” that two GR show the same fraction, rather than reasoning based on their constituent KCs (i.e., based on reasoning about the numerators and the denominators of the fraction shown in each GR being the same). Fluent learners can treat one GR as a single perceptual chunk, which allows them to perform quickly and effortlessly with MGRs. Hence, fluency in making connections between MGRs allows students to “simply see” that different GR show the same fraction, without having to reason about their equivalence based on corresponding KCs. The top part of Fig. 3 depicts this process in making connections between MGRs of fractions.

But, how do these – as of now separate – theoretical models for sense-making processes and fluency-building processes fit together? I propose to investigate this question with regard to connection making between MGRs of fractions. Before describing the proposed study in detail, I will describe the prior work I have done as part of my dissertation project which provides empirical support for the theoretical model just discussed.

3 Prior Work
I have conducted several classroom experiments which have investigated (1) whether the processes described as part of the theoretical model play a role in students’ robust learning of fractions, and (2) how we can support these processes within an ITS.

3.1 Experiment 1: Self-explanation prompts as support for understanding of individual GRs
Research shows that dual representations can significantly enhance students’ learning: students typically learn better from a combination of text and graphics than from text alone. Furthermore, there is evidence that the positive effect of learning with dual representations is mediated by an increased engagement in self-explanation activities (i.e., the process of generating explanations to oneself with the goal to make sense of what one is learning (Chi, Bassok, Lewis, Reimann, & Glaser, 1989): students who generate more high-quality self-explanations also show the highest learning gains when working with dual representations. Ainsworth and Loizou (2003) hypothesize that dual representations are beneficial because it can promote the self-explanation effect. Berthold and colleagues (2008) prompted students to self-explain while studying multi-
representational worked examples (i.e., instructional examples in which each step of the correct solution is provided). They found that prompting learners to relate different representations (textual descriptions, one GR, and symbols) promoted conceptual and procedural knowledge.

As part of my diploma thesis (the German equivalent to a Master thesis), I conducted an experiment to investigate whether the advantage of dual representations generalizes the more complex, multi-representational learning materials often found in real educational settings which include multiple *graphical* representations. Furthermore, Experiment 1 aimed at investigating whether students’ benefits from MGRs depends on being prompted to self-explain the relation between each GR and the symbolic representation. One-hundred thirty-two 6th-grade students worked with one of four versions of an ITS for fractions for 2.5 hours during their regular math instruction. The versions of the Fractions Tutor varied on two experimental factors: number of representations (a single GR or MGRs) and self-explanation prompts (with or without prompts). The self-explanation prompts were designed to help students relate the GR to the symbolic notation while emphasizing KCs such as numerator and denominator.

Results based on pretest, immediate and delayed posttest scores from 112 students demonstrated no main effect of number of GRs, and a main effect of self-explanation prompts only on reproduction of conceptual knowledge. However, there was a significant interaction between number of GRs and self-explanation prompts on reproduction of conceptual knowledge, and on transfer of procedural knowledge at the immediate posttest: students in the prompted conditions performed better with MGRs, whereas students within the no-prompt conditions performed worse when provided with MGRs, compared to learning with a single GR. Please refer to Rau, Aleven, and Rummel (2009) for a more detailed description of Experiment 1.

Although Experiment 1 was part of my work prior to the Ph.D., it laid the foundation for the experiments I have conducted as part of my dissertation project, described in the following. Experiment 1 demonstrates that *understanding* of each individual GR in terms of its KCs is an important prerequisite for the effectiveness of MGRs. Furthermore, self-explanation prompts were shown to be a successful means to support students in relating GRs and symbolic representations based at the level of corresponding KCs. This type of sense-making support enabled students to benefit from MGRs in acquiring robust knowledge about fractions.

3.2 Experiment 2: Interleaving task types as support for understanding of individual GRs
In areas where learners engage in extended problem-solving practice with MGRs across several task types, instructors and instructional designers must decide how to sequence GRs (e.g., circle and number lines) and task types (e.g., finding equivalent fractions and comparing fractions). Should they interleave MGRs while blocking task types, or should they interleave task types while blocking MGRs? What sequence will lead to the most robust learning gains?

The literature on contextual interference gives reason to believe that the decision of whether to interleave MGRs or task types will influence students’ learning. Generally, the results from contextual interference research have demonstrated that “interleaved practice” leads to better learning results than “blocked practice” (Battig, 1972; Schmidt & Bjork, 1992). In this research, the independent variable has typically been whether learning tasks were presented in
“blocks” of the same type (e.g., task 1 – task 1 – task 1 – task 2 – task 2 – task 2 – task 3 – task 3), or whether learning tasks of different types were interleaved (e.g., task 1 – task 2 – task 3 – task 1 – task 2 – task 3 – task 1 – task 2 – task 3). The contextual interference effect has been demonstrated in a variety of domains including algebra (e.g., Taylor & Rohrer, 2009), troubleshooting (e.g., de Croock, Van Merrienboer, & Paas, 1998) and decision-making tasks (Helsdingen, van Gog, & Van Merrienboer, 2011). However, the research on interleaved practice has not investigated whether the dimension on which learning tasks are interleaved (e.g., MGRs or task types) matter. In other words, it remains an open question whether MGRs are more effective when they are interleaved, or when task types are interleaved.

To answer this question, Experiment 2 manipulated the dimension on which tutor problems were interleaved: students either switched frequently between task types but infrequently between MGRs (e.g., A1 – A2 – A3 – B1 – B2 – B3 – C1 – C2 – C3, where A, B, and C are different GRs, and the numbers are different task types), or they switched frequently between MGRs but infrequently between task types (e.g., A1 – B1 – C1 – A2 – B2 – C2 – A3 – B3 – C3). The study involved 269 students in grades 5 and 6 who worked with one of two versions of the Fractions Tutor for 5h during their regular math class.

An analysis of the scores from 158 students on a pretest, an immediate and a delayed posttest a significant main effect for condition: interleaving task types while blocking MGRs lead to better conceptual knowledge of fractions representations (including symbolic representations) than interleaving MGRs while blocking task types (Rau et al., 2010; Rau et al., in press). I argue that task types (e.g., equivalent fractions and fraction comparison) are more variable than the GRs used in the Fractions Tutor (e.g., circles and number lines). Task types explicitly require students to apply different operations (such as finding equivalent fractions or comparing fractions). By contrast, GRs implicitly provide different conceptual views on the problem at hand (by depicting a fraction as a shaded part of a circle, or as a dot on a number line), and the conceptual differences (i.e., fractions as parts of a whole, or fractions as a measure) might be difficult to discern for novice learners. In fact, the GRs are designed to be intuitive: they employ perceptual processes in an effective and easy-to-understand way. In order to use GRs, students are not expected to engage in explicit reasoning about the properties of the representations. It thus seems reasonable to assume that the conceptual differences between the GRs were not as salient as the differences between task types. Interleaving learning tasks along the dimension of greatest variability might most effective because greater problem variability is likely to increase the need for repeated reactivation of knowledge. Reactivation of task-specific knowledge while gaining extensive practice with individual GRs across a sequence of task types promoted students’ conceptual understanding of fractions representations. Taken together, Experiment 2 provides evidence that interleaving task types rather than interleaving GRs supports students’ understanding of individual GRs, which lead to robust learning of fractions.

3.3 Experiment 3: Interleaving GRs as support for fluency with individual GRs
While Experiment 2 investigated the effects of interleaving learning tasks on one dimension (i.e., task types) compared to interleaving them on another dimension (i.e., GRs), the question of
whether interleaving GRs promotes students’ learning when using a constant sequence of task types remained open. Experiment 3 was designed to address this question: when moderately interleaving task types, how frequently should instruction switch between different GRs? Blocked practice with MGRs may be successful in promoting understanding of GRs as students can get in-depth experience in using one GR across a sequence task types. In fact, Experiment 2 showed that blocked practice with MGRs along with interleaved practice with task types lead to better conceptual knowledge of fractions representations. Interleaving MGRs, on the other hand, might allow students to frequently reactivate their knowledge about representations, which should strengthen their memory and increase the likelihood that students can fast and effortlessly use their knowledge about individual GRs in subsequent learning tasks. In other words, interleaved practice might increase students’ fluency with individual GRs. If fluency with GRs builds on understanding of individual GRs, then students may benefit most from a condition that gradually moves from blocking to an increasingly interleaved sequence of GRs.

I investigated this question in a classroom experiment with 587 4th- and 5th-grade students from three school districts. Students worked with one of four versions of the Fractions Tutor in which MGRs (circles, rectangles, and number lines) were either blocked, moderately interleaved, fully interleaved, or increasingly interleaved. In an additional control condition, students worked with only a single GR (i.e., only circles, only rectangles, or only number lines).

Based on pretest, an immediate and a delayed posttest scores from 290 students, I found a significant advantage of the fully interleaved condition compared to the blocked, the moderately interleaved, and the increasingly interleaved conditions on conceptual transfer at the delayed posttest, especially for low prior knowledge students. Furthermore, there was a marginally significant advantage for the increasingly interleaved condition over the blocked, moderately interleaved, and fully interleaved conditions on fluency with the number line at the immediate and the delayed posttests. Although a comparison of all MGR conditions and the single-GR condition was not significant for the majority of dependent measures, the interleaved MGR condition significantly outperformed the single-GR conditions on fluency with the number line, conceptual transfer, and marginally on procedural transfer.

Taken together, the results from Experiment 3 demonstrate that interleaved practice with MGRs (in addition to moderately interleaving task types) leads to better learning of fractions than blocked practice. Furthermore, only when provided in an interleaved fashion were MGRs more effective in promoting learning of fractions than a single GR. But how do we know that this effect is not due to students making connections between MGRs across consecutive tutor problems – which would indicate that interleaved practice with MGRs supports students ability to make connections between MGRs rather than their knowledge about individual GRs?

To address this question, I conducted a small-scale think-aloud study with six students who worked with the fully interleaved version of the Fractions Tutor. The goal of the think-aloud study was to assess what kinds of spontaneous connections students make between MGRs across consecutive tutor problems, and whether students’ ability to make these connections can be enhanced by prompting them to do so. Results showed that students made almost no spontaneous
connections between MGRs: when students were not prompted to make connections, only five utterances were coded as connections. When prompted to compare the MGRs, however, students were able to generate connections. (Also see Rau, Rummel, et al., 2012.)

Based on the results from the think-aloud study, it seems likely that the advantage of interleaved practice in Experiment 3 is not due to students’ conscious connection making between MGRs, but from repeated reactivation of knowledge about GRs, which, as argued, might help students in developing fluency with individual GRs.

The finding from Experiment 3, that one should interleave practice with MGRs, do not contradict the finding from Experiment 2, that one should interleave task types rather than MGRs. Experiment 3 built on the outcomes of Experiment 2 in that a moderately interleaved sequence of task types was used consistently across conditions. Experiment 3 therefore answered the question whether in addition to interleaving task types, MGRs should also be interleaved. Furthermore, interleaved sequences of MGRs and task types might serve different purposes: interleaving task types serves the development of understanding of individual GRs, whereas interleaving MGRs serves the development of fluency with individual GRs. At different times during the learning process, different relative sequences of MGRs and task types might be most beneficial to students’ learning of the domain. The relatively good performance of the increasingly interleaved condition in Experiment 3, which gradually moved from a blocked sequence to a more and more interleaved sequence of MGRs, speaks to this hypothesis: blocking MGRs while interleaving task types might be most beneficial early in the learning sequence, whereas interleaving MGRs (in addition to interleaving task types) might become more important later during the learning process.

3.4 Experiment 4: Supporting understanding and fluency in representational flexibility
The goal of Experiment 4 was to investigate the complementary roles of ITS support for sense-making processes and for fluency-building processes in connection making between MGRs. As this experiment is central to the study I propose to conduct as part of my dissertation, I will describe Experiment 4 in more detail.

3.4.1 Theoretical background and hypotheses
Most research on connection making has focused on the support of either process, rather than on supporting both. It seems likely, however, that in complex domains both learning processes, sense-making processes and fluency-building processes play a role (Koedinger, et al., in press).

Research on sense-making processes in connection making have typically investigated ways to help students relate corresponding elements of domain-relevant concepts (Bodemer & Faust, 2006; Bodemer, et al., 2004; Brünken, et al., 2005; Seufert & Brünken, 2006; van der Meij & de Jong, 2006). These studies have typically demonstrated sense-making processes in making connections are crucial for students’ acquisition of domain knowledge.

Research on fluency-building processes in connection making between have investigated the effects of perceptual training in relating representations on students’ math learning (Kellman & Garrigan, 2009; Kellman, et al., 2008; Kellman, Massey, & Son, 2009). Students learned to find corresponding representations of math problems, such as textual descriptions, GRs, and
symbolic representations. Fluency training aims at providing students with experience without asking them to consciously reflect on the learning content. Rather, the training helps students more efficient at extracting structurally relevant information across a variety of representations through experience and discovery. Results of these studies show that students who already have a good conceptual understanding of the domain and of the representations tend to perform badly on fluency tests, which assess the accuracy with which students recognize and construct corresponding representations (Kellman, et al., 2008; Kellman, et al., 2009). Students who received fluency training subsequently performed better not only on fluency tests, but also on tests of conceptual and procedural knowledge, compared to students who did not receive such training (Kellman, et al., 2009). Effects of fluency training were demonstrated both for experts (i.e., students with relatively good conceptual understanding of the domain) and novices in algebra and fractions learning.

Experiment 4 was designed to investigate the hypothesis that students will learn best about fractions when they receive support for both sense-making processes and fluency-building processes in making connections between MGRs. A further goal of Experiment 4 was to investigate how best to support students in making sense of connections between MGRs. Specifically, Experiment 4 investigated how much automated support students should receive from the system. On the one hand, providing students with auto-linked GRs (i.e., GRs that are linked in such a way that the student’s manipulations of one are automatically reflected in the other) has been shown to lead to learning in complex domains (van der Meij & de Jong, 2006). On the other hand, research has demonstrated that students should actively create connections between representations, rather than passively observing correspondences (Bodemer & Faust, 2006; Bodemer, et al., 2004; Gutwill, et al., 1999). I thus compared two types of sense-making support, one in which the tutor demonstrates connections (i.e., auto-linked representations), and one in which that burden falls on the student. A well-researched way of supporting sense-making processes is to provide students with worked examples; an instructional intervention that has been shown to be effective in many domains (Renkl, 2005). Berthold and Renkl (2009) compared students’ learning from multi-representational worked examples to single-representation worked examples and found that multiple representations can enhance students’ learning from worked examples. However, worked examples have not yet been investigated as a means of support for connection making between multiple representations. In Experiment 4, worked examples were used which provided a more familiar GR as a guide to solve an isomorphic problem that involved a less familiar GR. Students were prompted to integrate the example problem and the new problem by making connections between the two GRs. I hypothesized that worked examples support (compared to auto-linking support) would be more effective in promoting students’ learning of fractions, since students have to engage actively in connection making.

A total of 1308 4th- and 5th-grade students participated in the experiment during their regular math class. Students either received sense-making support for connection making (in the form of either auto-linked [AL] support or worked example [WE] support) or not. This factor
was crossed with a second experimental factor, namely, whether or not students received fluency-building (FB) support for connection making or not. Since many education researchers and practitioners emphasize the importance of number lines (NMAP, 2008; Siegler et al., 2010), an additional control condition was implemented which used only the number line.

3.4.2 Worked examples as sense-making support for connection making

Fig. 4 shows the screen shot of a sense-making problem for equivalent fractions. Students learn that equivalent fractions in different GRs show the same amounts or lengths that are cut into different numbers of sections. They also learn that numerators and denominators of equivalent fractions are always expanded by the same multiplier. Students are first presented with the worked example (part A in Fig. 4). To ensure that students read through the worked examples, they are asked to fill in the last step themselves (step A-3 in Fig. 4). Once they complete that step, the problem-solving part of the worked example appears on the right (part B in Fig. 4). The side-by-side arrangement between corresponding steps in the worked example and the problem was chosen to assist students in aligning corresponding aspects of the worked example and the problem. At the end of each worked example problem, students receive reflection prompts that help them abstract a general principle from the two GRs (part C in Fig. 4). For each step, students receive feedback at the level of the relevant KCs of the step at hand, for instance, by explaining that the number of total sections a rectangle is cut into correspond to the denominator.

Fig. 5 shows a screen shot of a sense-making problem for fraction comparison. Sense-making processes of connection making are being supported by demonstrating that fractions can be judged based on their relative size to one another. The Fractions Tutor focuses on the concept of inverse relationships between the number of total sections and the size of each section. The setup of the problem (i.e., completion of last step in the worked example, alignment of corresponding steps in worked example and problem, and reflection prompts) corresponds to that support described for the equivalent fractions example (Fig. 4).

3.4.3 Mixed representations as fluency-building support for connection making

Fig. 6 shows a fluency-building problem for equivalent fractions. Here, students sort a variety of equivalent GRs using drag-and-drop. Rather than identifying numerator and denominator to solve the equivalence problem computationally, students visually judge whether GRs show equivalent fractions by estimating their relative size. Feedback is given only about the correctness of the sorting task, without mentioning the KCs of numerator and denominator.

Fig. 7 shows an example of a fluency-building problem for fraction comparison. Students sort GRs based on their relative size, using drag-and-drop. Again, students are encouraged to visually estimate the relative size of a variety of GRs.

3.4.4 Results

Due to data loss, only one school district (N = 599) had complete data. The results from this school district showed no significant main effect for sense-making support or fluency-building support, but a significant interaction between sense-making support and fluency-building support on conceptual knowledge: students who received both types of support performed best on a conceptual knowledge posttest. Further, students who received WE and FB support significantly
outperformed students who received only FB support and students who received AL support and FB support. Finally, although there was no overall advantage compared to the number-line condition, the WE-FB condition significantly outperformed the number-line condition on conceptual knowledge. (Also see Rau, Aleven, et al., 2012.)

Taken together, Experiment 4 provides evidence that both sense-making support and fluency-building support for connection making are needed in order to promote students’ conceptual learning of fractions. Furthermore, the results demonstrate that students need to actively engage in sense-making activities. A novel application of WEs is a successful means to support students in actively making these connections. Finally, the fact that only the WE-FB condition significantly outperformed the number-line condition demonstrates that students’ benefit from MGRs builds both on sense-making processes and fluency-building processes.
3.4.5 Auxiliary analyses

Although Experiment 4 integrates two different, thus far separate lines research on sense-making processes and fluency-building processes in connection making, it raises interesting new questions about the relation between these learning processes. It is surprising that there were no significant main effects for sense-making support or fluency-building support; only the combination of both was effective. Did one type of support enable students to benefit from the other? To develop hypotheses for this question, I conducted an analysis of the errors students made while working on the Fractions Tutor. Specifically, I was interested in the types of errors that students in the WE, FB, and WE-FB conditions made on the worked example problems and on the fluency-building problems. I compared the frequency of error types on those connection-making problems that were the same for two given conditions: errors on worked examples problems in the WE and the WE-FB conditions, and errors on fluency-building problems in the FB and the WE-FB conditions. Based on Chi-square tests and regression analyses, I determined which error types differed significantly between these conditions, and which were significant predictors of students’ performance on the conceptual posttests. I then included these error types in a structural equation model (SEM) to investigate potential mediators of the combined effect of sense-making and fluency-building support.

Fig. 8 shows the results from the SEM for the comparison of the WE and the WE-FB conditions based on errors students made on the worked example problems. Students in the WE-FB condition, compared to the WE condition, make more SE errors (i.e., errors in answering self-explanation prompts) and more place1 errors (i.e., errors in finding 1 on an unlabeled number line), both of which decreased performance on the conceptual posttest. Fluency-building support has a direct positive effect on posttest performance which is stronger than the sum of the negative mediation effects. While the SEM for the WE and WE-FB conditions provides insights into potential costs of fluency-building support, it does not help identify mediators of the positive effect of fluency-building support provided in addition to sense-making support.

Fig. 9 shows the best fitting SEM for the FB and WE-FB conditions based on the errors students made on fluency-building problems. Students in the WE-FB condition make more nameCircleMixed errors (i.e., errors in identifying the fraction depicted by a circle), but fewer improperMixed errors (i.e., errors in identifying an improper fraction) and equivalence errors (i.e., errors in identifying equivalent fractions) than students in the FB condition. Students who make fewer nameCircleMixed errors also make more subtractionMixed (i.e., errors in finding the difference between two given fractions) and improperMixed errors, which decrease performance in the conceptual posttest. These mediations demonstrate a negative effect of sense-making support (provided in addition to fluency-building support) on conceptual posttest performance, while controlling for pretest performance. However, sense-making support also has a positive effect on posttest performance, mediated by fewer improperMixed errors and equivalence errors.

The findings from the SEM demonstrate that sense-making support reduces certain error types made on fluency-building problems, thereby leading to better performance on the conceptual posttest. However, I did not find evidence of an advantage of fluency-building
support based on errors made on sense-making problems which were associated with higher posttest performance. The SEM analysis thus leads to the hypothesis that sense-making support helps students benefit from fluency-building problems by reducing certain types of errors on fluency-building problems. However, the SEM analysis does not support the notion that fluency-building support helps students benefit from sense-making problems.

4 The Fractions Tutor

In addition to developing a theoretical model and instructional principals for designing support for learning with MGRs, a further outcome of my PhD research is an ITS for fractions. In this section, I will describe the processes involved in the iterative development of the Fractions Tutor and its impact on students’ robust learning.

4.1 User-centered identifications of misconceptions of the Fractions Tutor

In order to help students learn, the Fractions Tutor needs to address students’ misconceptions about fractions. Before designing new tutor problems, I conducted interviews with students from the target population (i.e., students from grades 4-6) using open-ended items to identify misconceptions about fractions and GRs. I also explored which GRs students drew on to explain their reasoning. Furthermore, I interviewed experienced math teachers about students’ difficulties with various topics, and which GRs they deem suitable to remedy their difficulties. Based on these interviews, I then identified task types and GRs that the Fractions Tutor should cover. The teacher interviews also showed that alignment with the Pennsylvania System of School Assessment (PSSA) is crucial to them as stakeholders. When reviewing existing materials for fractions instruction, I therefore also included PSSA materials to ensure alignment with the Fractions Tutor.

The actual tutor problems were developed based on a cognitive task analysis (e.g., Baker, Corbett, & Koedinger, 2007; Koedinger & Nathan, 2004) of the selected topics. The analysis
aimed at identifying the KCs that the selected topics involved, and that, consequently, the Fractions Tutor should explicitly address. I designed the tutor interfaces based on a sub-goaling strategy (e.g., Anzai & Simon, 1979, VanLehn, 1991) to provide explicit guidance with respect to the KCs involved in each step. The development of new tutor problems was an iterative process during which an experienced math teacher reviewed problems, suggested improvements, reviewed the revised problems, etc. Next, I created light-weight prototypes on paper pilot-tested them. Students were asked to think aloud as they solved the prototype problems, which allowed to evaluate whether the problem addressed students’ misconceptions.

Based on the results from the prototype pilots, I then created high-fidelity prototypes which I again evaluated and iteratively improved based on small-scale think-aloud studies. Finally, I created full-scale tutor problems and piloted them in the lab or at a school. Again, the goal was to evaluate how well the problems address misconceptions about fractions and GRs.

4.2 Iterative development based on classroom experiments
Development of the Fractions Tutor obviously did not stop after its deployment in classrooms. Based on interviews and surveys with teachers and students who used the Fractions Tutor, I identified issues (e.g., wording in hint messages) and resolved them. I also used the tutor log data to identify steps students struggled with while working with the Fractions Tutor. Based on pre- and posttest data, I identified topics on which students did not learn as much as expected. I then revisited these steps in a think-aloud study and worked together with an experienced math teacher to identify reasons for students’ difficulties with these steps. Often, this process led to reformulation of questions, hint messages, or error feedback messages. In some cases, I decided to restructure tutor problems. Such modifications were then tested in further think-aloud studies.

4.3 Interface design based on learners’ needs
One issue that repeatedly drew students’ and teachers’ attention during interviews were aesthetic aspects of the Fractions Tutor. Students commented that they would like the tutor to be more “flashy”, “exciting”, and more “colorful”. Teachers also often commented on the structure of the tutor problems. For instance, they felt that students often lost track of the overarching task in the tutor problem: it seemed that the tutor interface did not optimally reflect the sub-goaling strategy. To address these issues, the design of the tutor interfaces received specific attention. The goal of the design process was to make the Fractions Tutor look more appealing for children in grades 4-6. At the same time, the goal was for the interface to emphasize the sub-goal structure of the tutor problem through its layout.

The result of the design process was a design template that can be applied to all different types of tutor activities. The new design also includes a variety of flashy success messages at the end of each tutor problem. Examples of the new interface design are shown in Figures 4-7 in section 3.4. I piloted the new interface design prior to the deployment of the Fractions Tutor in classrooms. Student and teacher interviews following the classroom study showed that the issues of interface aesthetics and sub-goaling did not persist. Furthermore, students seemed to really like the success messages.
4.4 Impact of Fractions Tutor on students’ learning about fractions

The Fractions Tutor covers a comprehensive set of topics including naming fractions shown in GRs, making GRs of fractions reconstructing the unit of fractions, improper fractions, equivalent fractions, fraction comparison, addition and subtraction. The content represents at least 10h of supplementary instruction for students in grades 4 and 5.

Results from the classroom experiments described in section 3 demonstrate substantial learning gains from pretest to posttest, based on about 3,000 students who have participated in these classroom studies. After working for 10h on the Fractions Tutor, students in Experiment 4 scored on average 10% higher (d = .40) on an immediate posttest which involved standardized test items (e.g., PSSA items), and 15% higher (d = .60) on a delayed posttest compared to an equivalent pretest. On items involving a number line, students scored 11% higher on the delayed post-test than on the pre-test. On an efficiency test (which accounted for the speed with which students solved number line problems), we found large effect sizes in learning gains of d = 1.17 (immediate post-test) to d = 1.21 (delayed posttest). This level of improvement is important because the number line is challenging, yet is key to fractions learning (Siegler et al., 2010) and later learning of, for example, algebra (NMAP, 2008).

4.5 Summary: A successful interplay of methods

Taken together, the Fractions Tutor is the result of an interplay of a variety of methods coming from educational and experimental psychology, instructional design, ITS, and HCI. It has been shown significantly (and robustly) improve students’ knowledge of fractions with medium to large effect sizes and covers more than 10h of effective instruction on fractions. I used HCI methods to ensure that the Fractions Tutor addresses students’ and teachers’ needs in learning and instructing about fractions. I have iteratively improved the Fractions Tutor based on lab studies and classroom experiments. The design of the interface is both visually appealing and aligns with ITS principles such as explicit support of KCs based on sub-goaling. By applying educational psychology theory, I have been able to identify open issues regarding the design of effective instructional materials with MGRs. My classroom experiments have provided answers to some of these issues, which has lead to the formulation of instructional design principles that can guide the development of further instructional materials. The Fractions Tutor embodies the lessons learned from these experimental studies.

5 Proposed Experiment

Based on several experimental studies, I have developed a theoretical model for learning with MGRs. In order for students to benefit from MGRs, they need to develop understanding and fluency with individual GRs, they need to develop understanding of connections between MGRs, and they need to become fluent in making these connections. The fact that in Experiments 3 and 4, only the conditions in which students received support for these processes outperformed a single-GR condition demonstrates that the acquisition of these cognitive competencies is necessary for students’ success in learning from MGRs. The merit in providing such support lies
in enhancing their conceptual domain knowledge: in most of the experiments described above, differences were found on conceptual knowledge and on transfer of conceptual knowledge.

However, it remains unclear how sense-making processes and fluency-building processes interact. This question arises from the findings in Experiment 4, in which neither sense-making support for connection making nor fluency-building support alone were effective, but only both types of support together enhanced learning. Based on an SEM of students’ error types, one might hypothesize that sense-making support helps students benefit from fluency-building support, rather than the other way around. But the SEM analysis cannot conclusively answer this question: In Experiment 4, fluency-building support was provided after sense-making support. It is unclear whether fluency-building support might have helped students benefit from sense-making support if the two types of support had been provided in the opposite order. Further, due to the selective nature of the SEM analysis (i.e., selection of error types only if they significantly differed between conditions and predicted posttest performance), its merit can only be hypothesis generation, but not empirical evidence for (or against) a hypothesis.

The question of how sense-making processes and fluency-building processes interact is also of practical relevance. If understanding (acquired through sense-making support) enables students to benefit from subsequent fluency-building support, we should support sense-making processes before fluency-building processes (understanding-first hypothesis). If, on the other hand, fluency enables students to benefit from sense-making support, we should support fluency-building processes before sense-making processes (fluency-first hypothesis). Providing support for these learning processes in the optimal order should maximize students’ benefit from activities designed to support connection making, which is – as argued – a crucial competence that will enhance students’ robust domain knowledge.

By investigating how sense-making processes and fluency-building processes in connection making interact, the proposed research is a step towards closing the gap between studies that have exclusively on sense-making support and those that have focused solely on fluency-building support. In the following, I will describe theoretical perspectives pertaining to competing hypotheses on the question of which learning process to support first.

5.1 Theoretical perspectives and hypotheses
A variety of literatures acknowledge that both understanding and fluency are important aspects of robust knowledge within a domain. Theories of cognitive skill acquisition (e.g., Anderson, 1983; Koedinger et al., in press; Ohlsson, 2008) describe both sense-making processes and fluency-building processes as integral learning processes that students need to engage in to master a domain. And although many educational practice guides for mathematics education almost exclusively stress the importance of conceptual understanding (e.g., Siegler et al., 2010) – maybe in an effort to counteract the longlasting emphasis on procedural learning – they have recently put more emphasis on fluency as well. For example, the NMAP (2008) describes fluency in relating different fractions representations as one important foundation for later algebra learning. Unfortunately, however, neither of these literatures makes claims about dependences between sense-making processes and fluency-building processes, thus obviating the
need for an empirical investigation of this question. In the following, I will briefly summarize arguments that speak for the hypothesis that one might expect the most robust learning gains when supporting fluency-building processes before sense-making processes (H1: fluency-first hypothesis), and arguments for the opposite prediction, that instruction will be most efficient when support for sense-making processes is provided before supporting fluency-building processes (H2: understanding-first hypothesis). I will then discuss specific predictions made by each of these hypotheses and how I plan to assess their accuracy.

5.1.1 Fluency-first hypothesis

Students who have the opportunity to become fluent in making connections between MGRs by visually relating them may benefit from increased “cognitive head room” during subsequent learning tasks (Koedinger et al., in press), which might involve sense-making processes about the conceptual nature of the connections between MGRs.

Indeed, Kellman et al. (2009) argue that fluency results from automation of the perceptual task to make connections between different representations. This type of fluency is acquired through experience with a variety of representations without having to engage in sense-making processes about how corresponding KCs are depicted in the different GRs. Fluency training reduces cognitive load by automating the perceptual task, thereby freeing up cognitive resources for more complex learning tasks. If the perceptual task is not automated, it will unnecessarily take up cognitive resources which might be missing for the completion of a more complex task, such as the task to make sense of connections based on corresponding KCs of GRs. Is this the case, providing sense-making support before fluency-building support may enhance chances of cognitive overload while students work on sense-making problems, which is known to hamper learning (Chandler & Sweller, 1991).

5.1.2 Understanding-first hypothesis

Kellman et al.’s (2009) account relies on the assumption that fluency in making connections – the automation of perceptually relating MGRs – can be learned independently from understanding of these connections. However, this assumption might not be true: understanding the connections between MGRs might equip students with the knowledge they need in order to benefit from fluency-building support. If students do not know what aspects of different GRs correspond to one another, how should they know what to attend to while solving fluency-building problems? Not having this knowledge may lead to inefficient learning strategies, such as trial and error, which might impede students’ benefit from fluency-building support.

Indeed, the mathematics education literature seems (albeit not explicitly) to agree with this view. Education practice guides, such as the NCTM standards (2010), provide “checklists” of knowledge that students should have acquired by specific grade levels. Understanding of fractions representations is expected by the end of grade 5. The ability to efficiently work with fractions representations is expected later – not before the end of grade 8.

Instructional Design principles based on Cognitive Load theory provide a theoretical rationale for the understanding-first hypothesis. Fluency-building problems comprise perceptually rich learning tasks while providing minimal guidance for students to solve them.
Students are expected to learn (through discovery) to extract structurally relevant information from experience across a variety of representations. Kirschner, Sweller, and Clark (2011) vigorously argue that minimally-guided practice with information-rich problems may increase cognitive load during problem solving and may hamper learning. In vivid terms, they describe that “minimally-guided learning does not enhance student achievement any more than throwing a non-swimmer out of a boat in the middle of the lake supports learning to swim” (p. 4).

5.1.3 Specific predictions by the understanding-first hypothesis and the fluency-first hypothesis

Both the understanding-first hypotheses and the fluency-first hypotheses make predictions about process-level mechanisms such as confusion, conceptual processing, and attention efficiency. The fluency-first hypothesis predicts that students who receive sense-making support first should, due to cognitive overload, experience confusion and have trouble to efficiently direct their attention to KCs relevant to the current step while working on sense-making problems. Furthermore, cognitive overload will reduce the ability to conceptually process connections between MGRs; instead, students might focus on superficial (i.e., conceptually irrelevant) features of the GRs, such as their color. In fact, how should they know what to attend to? Finally, students are expected to make more errors while working on sense-making problems if they have not yet acquired fluency. By contrast, the understanding-first hypothesis predicts that students who receive fluency-building support first will show confusion and inefficient direction of visual attention while working on fluency-building problems, because these problems demand too much of them. Cognitive overload should also hamper students’ ability to conceptually process connections between MGRs. Finally, students should make more errors while working on fluency-building problems as they demand too much of students who have not yet acquired understanding of connections.

I propose to use eye tracking to investigate the role of visual attention, and of confusion to the extent that it impacts efficiency in visual attention direction. Specifically, I will assess the efficiency with which students direct their attention to relevant KCs (i.e., the “relevant” area of interest [AOI]: such as the colored sections of a circle diagram when asked for the numerator of a fraction) while solving sense-making problems and fluency-building problems. Confusion should manifest itself in longer time to first fixation of the relevant KC of the current problem-solving step as students experience difficulties in identifying the relevant aspects of the GR. Efficient visual attention should lead to longer fixation of the relevant AOI, whereas confusion is expected to lead to shorter fixation lengths across a variety of AOIs, not only of the relevant AOI, but also of AOIs that are irrelevant for the current step. The visual-attention patterns just described should coincide with lower performance in solving tutor problems (e.g., error-rate as identified from the tutor log data).

To get insights into how students make connections between MGRs, I propose to collect cued retrospective reports (CRRs). Of interest is whether students make surface-level connections or conceptual connections between MGRs. Gaining insights into the nature of students’ connection making can disambiguate patterns that eye-tracking data cannot distinguish. For instance, a student may frequently switch between relevant AOIs in circle and number line...
(suppose the relevant AOI for a given step corresponds to the numerator of a given fraction) and attend to conceptually relevant aspects of the GRs (e.g., they both show a numerator of 3), or to irrelevant aspects of the GRs (e.g., they have the same color). This example illustrates the purpose of collecting verbal data: the goal is to gain insights into why students pay attention to a certain aspect of the tutor problem, and how they decided to attend to these aspects. The commonly used procedure of asking students to think aloud while solving a problem, by verbalizing any thought that comes to their mind, may not be a suitable method to learn about the nature of students’ connection making. However, think-alouds tend to yield information about actions and outcomes of problem solving (Taylor & Dionne, 2000), rather than information about strategies or metacognitive knowledge (Van Gog et al., 2005). Retrospective reporting has been shown to be more suitable at providing information about the why and how of learning processes (Taylor & Dionne, 2000; Van Gog et al., 2005). A disadvantage, however, of retrospective reporting is that it relies on participants’ memory of why and how they solved a problem. By contrast, think-aloud protocols have the advantage that they are collected concurrently with students’ problem solving. CRRs seek to combine the advantages of think-aloud protocols and retrospective reporting by using records of eye movements superimposed on a record of students’ problem-solving behaviors to cue students to verbalize their thought processes. Thus, I propose to use CRRs by showing students a record of their eye movements superimposed on a video of their interactions with the tutor problem. Students will be asked to describe how they solved the connection making problems. I will use this data to identify differences between conditions with regard to surface-level and conceptual connections.

5.2 Experimental design and procedure
To answer the question of how sense-making processes and fluency-building processes interact, I propose to conduct a lab experiment with two experimental groups. Students will be randomly assigned to either the sense-making support before fluency-building support condition (S-F) or to the fluency-building support before sense-making support condition (F-S).

Students in the S-F condition will work on four sense-making problems for one task type covered by the Fractions Tutor. Next, they will work on four fluency-building problems for the same task type. For another, unrelated task type, students will then work on four sense-making problems, followed by four fluency-building problems for the same task type. The F-S condition will follow the same procedure, except that they will receive fluency-building problems before sense-making problems. Students’ knowledge of fractions will be assessed before and after their work with the Fractions Tutor. Furthermore, I will measure of understanding and fluency after students’ completed one round of sense-making support or fluency-building support. This manipulation check serves to assess whether the sense-making problems enhance only understanding (but not fluency), and whether the fluency-building problems promote only fluency (but not understanding). As CRRs might enhance students’ learning (since revisiting the tutor problems will expose them to the learning content again), I propose to collect CRRs after the posttest. I will randomly select two tutor problems of each support type / task type
combination for each participant as cue for the CRRs. Tables 1 illustrates the sequence of activities and measures for the S-F condition.

Table 1. Sequence of activities and measures for the S-F condition, measures in italic font.

<table>
<thead>
<tr>
<th>Pretest</th>
<th>Task type 1: 4 x sense-making support</th>
<th>Test of understanding / fluency</th>
<th>Task type 1: 4x Fluency-building support</th>
<th>Test of understanding / fluency</th>
<th>Posttest</th>
<th>Cued retrospective reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task type 2:</td>
<td>4 x sense-making support</td>
<td>Test of understanding / fluency</td>
<td>Task type 2: 4x Fluency-building support</td>
<td>Test of understanding / fluency</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.3 Measures
A variety of measures will be employed in the proposed experiment: eye-tracking measures, CRR measures, manipulation check measures, and pre- and posttest measures. In the following, I will describe how I plan to compute each of these measures.

5.3.1 Eye-tracking measures
As mentioned, I propose to assess efficiency of visual attention direction in several ways. First, I will compute the time to the first fixation of the relevant aspect of the problem-solving step at hand. To this end, I will identify AOIs for each problem step that correspond to relevant KCs. For instance, when a student is explicitly instructed to relate circle and number line, relevant KCs are numerator and denominator. For each KC, instances can be identified for the circle, the number line, and the symbolic fraction; and each of these instances will correspond to an AOI. In step 1, students may be asked to identify the numerator, and in step 2, to identify the denominator. For each step then, I will annotate AOIs with the labels “relevant” and “irrelevant”: in step 1, numerator AOIs are relevant, but denominator AOIs are irrelevant. The variable time to first fixation of relevant AOI (time-firstRelevant) will be computed as the time it took students until they first fixated on one of the AOIs that are relevant to the problem-solving step at hand.

Second, I will assess the total fixation duration on the relevant AOIs. Again, I will compute this measure for each step. For example, while working on the numerator step, a student might look from the circle-numerator AOI to the number-line-numerator AOI. The variable duration of fixation on relevant AOIs (durationRelevant) will correspond to the sum of the fixation durations on AOIs that are relevant to the problem-solving step at hand.

Third, I will look at the frequency of switches between relevant AOIs. This variable will disambiguate two scenarios that durationRelevant cannot distinguish: (1) a student fixates on one relevant AOI for a very long time, or (2) a student switches between several relevant AOIs. Scenario 1 may indicate boredom or confusion rather than connection making. On the other hand, the student in scenario 2 appears to engage in connection making between relevant KCs of different GRs. The variable switches between relevant AOIs (switchRelevant) will be computed as the number of switches between AOIs that are relevant to the problem-solving step at hand.
Finally, I will compute the frequency of switches between irrelevant AOIs. A student who switches frequently between irrelevant AOIs or between relevant and irrelevant AOIs might be confused about how to solve a problem. The variable switches between irrelevant AOIs (switchIrrelevant) will be computed as the number of switches that move through AOIs that are irrelevant or relevant to the problem-solving step at hand.

5.3.2  Cued retrospective report measures
The protocols obtained from CRRs will be coded for conceptual processing and surface-level processing of connections. I will code each connection made between MGRs based on the coding scheme used in Experiment 3 (see Rau, Rummel, et al., 2012). The variables concConnect and surfaceConnect will correspond to absolute number of connections coded as conceptual and as surface connections, respectively. Since students may differ in terms of their ability to verbalize their learning process, I plan to Z-standardize concConnect and surfaceConnect.

5.3.3  Tutor log data
The Fractions Tutor logs all of the students’ interactions during problem solving. Of interest is how many errors students make as they work on the connection making tasks. I will compute error-rate as the number of errors per first attempt.

5.3.4  Manipulation check measures
The manipulation check measures serve to assess students’ acquisition of understanding and fluency. The test of understanding will include items that directly correspond to the tutor problems designed to support sense-making processes. To check students’ fluency, they will be asked to solve problems that correspond to the fluency-building problems as fast as they can.

5.3.5  Pre- and posttest measures
The pre- and posttest measures will serve to assess students’ knowledge of fractions, of individual GRs, and of connections between MGRs. I will adapt the test used in Experiment 4 to the choice of task types for this planned experiment. The test items will include reproduction and transfer items, as well as adapted standardized test items.

5.4  Materials
All students will work on the same tutor problems, provided in different order. The tutor problems I propose to use for the planned experiment were described in sections 3.4.2 and 3.4.3. A few modifications are necessary with respect to the tutor interfaces. I plan to adjust the size of the interfaces to match the screen resolution of the eye tracker. Furthermore, I will modify the interfaces by reducing “white spaces” (i.e., areas not covered by GRs, symbolic notations, or textual descriptions), and enlarging text areas and GRs to maximize the size of the AOIs.

5.5  Planned analyses
Several sources of data will be available for analysis: process-level data (i.e., eye-tracking data, cued retrospective report data, tutor log data), manipulation check data, and pre- and posttest data. In the following, I will describe which analyses I plan to carry out on these data sources.
5.5.1 Analysis of process-level data: Eye-tracking data, CRR data, and tutor log data

All process-level data should be analyzed in conjunction. The dependent variables are expected to be statistically related because all of them originate from the same students working on the same tutor problems. Further, I expect variance in the dependent variables to be partially explained by students’ pretest score. Therefore, I plan to use a MANCOVA (multivariate analysis of covariance) with condition (i.e., S-F vs. F-S) as between-subjects factor, support type (i.e., sense-making support and fluency-building support) as within-subjects factor, pretest score as covariate, and the available process-level measures as dependent measures.

A power analysis to assess the number of students $N$ needed for a MANCOVA with 2 between-subjects groups, 2 within-subjects groups based on an α-error probability of 0.05 and a power β of 0.8 demonstrated that an $N$ of at least 57 students is needed to detect a medium effect size of $f^2 = 0.15$ in within-between factors interactions. However, since (Holmqvist, et al., 2011) suggest that one should expect attrition of 20% of the data when using eye tracking due to calibration errors or other data quality concerns, I plan to collect data from at least 72 students.

5.5.2 Analysis of manipulation-check data

Two variables are available from the manipulation-check data: the score on the understanding test and the score on the fluency test. To test the understanding-first and the fluency-first hypotheses, I will carry out a MANCOVA with condition as a between-subjects factor, period as within-subjects factor, pretest score as covariate, understanding-score and fluency-score as dependent measures. A power analysis on a MANCOVA with 2 between-subjects groups, 2 within-subjects groups based on an α-error probability of 0.05 and a power β of 0.8 to assess the minimum effect size that can be detected with $N = 72$ students demonstrated that a small to medium effect size of $f^2 = 0.05$ in within-between factors interactions is detectable.

It will also be interesting to further investigate properties of the understanding and fluency assessments. I plan to conduct a confirmatory factor analysis on understanding and fluency items, to test the assumption that understanding items and fluency items test separate skills, and whether these skills differ by topic. To this end, I will contrast a general model (all items load on only one factor), a task-type model (two factors: equivalence and comparison), a learning-process model (two factors: understanding and fluency), and a combined model (four factors: understanding-equivalence, understanding-comparison, fluency-equivalence, fluency-comparison). I will also explore difficulty and discrimination properties of test items by applying an IRT model I have used for the test of fractions knowledge used in my previous experiments.

5.5.3 Analysis of pre- and posttest data

To analyze the effect of condition on students’ learning of fractions, I will use an ANCOVA with condition as the independent factor, pretest score as the covariate, and posttest score as the dependent measure. A power analysis demonstrated that an ANCOVA with $N = 72$ students can detect a small to medium effect size of $f^2 = 0.05$ in within-between factors interactions.

5.5.4 Causal path model

A suitable method for analyzing the causal relationship between the outcome measures and process-level measures is SEM. A hypothesized model with edges from pretest to posttest, from pretest to all seven process-level measures, from condition to all process-level measures, and
from all process-level measures to posttest has 23 degrees of freedom. A power analysis based
on McCallum, Browne, and Sugawara’s (1996) method of close fit determined that a sample size
of \( N = 166 \) is needed to verify the accuracy of this model using the RMSEA statistic for model
fit, assuming a critical RMSEA of 0.05. Unfortunately, it is unlikely that a lab experiment will
yield a sample of \( N = 166 \). For this reason, I plan to include only those process-level variables
that differ significantly between conditions. I will test the hypothesis model and compare it to
other plausible models, identified with Tetrad’s search function.

5.6 Time line
The activities necessary to complete the proposed research fall into four categories. First, several
preparatory activities are necessary to ensure the success of the planned experiment. As
mentioned, the tutor interfaces need to be adapted for the purpose of eye tracking. During this
phase, I also plan to conduct two pilot studies. I will conduct think-alouds combined (if needed)
with interviews about the aspects of the tutor interface they attend to while solving the tutor
steps. This pilot study will help adjust the hypotheses described in section 5.5.1 as to which
AOIs are relevant or irrelevant for each tutor step. Based on this first pilot study, I will create
AOIs and annotate them with dynamic labels (“relevant” or “irrelevant”) for each step. (This
pilot study is already under way.) A second pilot study will help to calibrate the position of
AOIs. To evaluate whether the relevant AOIs accurately reflect that successful learners attend to
them, and whether unsuccessful learners attend to irrelevant AOIs, I will collect CRRs after each
tutor problem. Participant recruitment will also occur as part of the preparatory phase.

Next, I will carry out the experiment, with the help of one or more research assistants. The
experimental phase can co-occur with some of the data-preparation and analysis phase,
which includes coding of CRRs, computation of process variables and outcome variables.

Once the analysis phase is completed, I will work on presentation activities. These
include interpreting the data, writing papers, writing the thesis statement, preparing the defense
talk, and the defense itself. Table 2 provides a more detailed time line for the planned activities.

Not listed in Table 2 are several auxiliary activities which will ensure my academic
progress beyond the PhD. I plan to submit two journal articles before completing my dissertation
as well as at least two conference papers. In the fall semester, I will also complete my final TA
requirement. Then, I plan to apply for post-doc positions starting in December.

<table>
<thead>
<tr>
<th>Week</th>
<th>Preparation activities</th>
<th>Experiment</th>
<th>Data prep &amp; analysis</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>8/20</td>
<td>Adapt tutor problems, recruitment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8/27</td>
<td>Adapt tutor problems, recruitment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9/3</td>
<td>Pilot 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9/10</td>
<td>Pilot 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9/17</td>
<td>Implement AOIs, recruitment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9/24</td>
<td>Implement AOIs, recruitment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/1</td>
<td>Pilot 2, recruitment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/8</td>
<td>Pilot 2, recruitment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/15</td>
<td>Make changes to AOIs, repeat pilot 2,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>recruitment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/22</td>
<td>Repeat pilot 2, recruitment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/29</td>
<td>Final fixes, recruitment</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6 Conclusion and expected contributions

Taken together, my dissertation will contribute to several strands of research. First, it will contribute to the vast literature on learning with multiple representations by providing an empirically tested theoretical model for learning with multiple graphical representations. This model extends existing models that have focused on dual representations (i.e., text and one GR) by making predictions for the specific, but (in educational practice) frequent case that students are presented with a variety of GRs each of which emphasizes a different conceptual aspect of a complex and challenging domain. This model has the potential to stimulate further research that can validate it and further refine it, and verify its applicability to other STEM domains.

My work also contributes to instructional design as it leads to concrete principles about how to present MGRs in real instructional materials. Though more research is needed to verify the generalizability of the instructional principles resulting from my work to materials other than intelligent tutoring systems (such as paper-based curricula), I see no reason to assume that they would not apply to other instructional materials as well.

My dissertation further contributes to cognitive science research as it integrates research that has been done separately on the development of understanding and fluency. Though the community has long since emphasized that both understanding and fluency are needed for mastery, little is known about the relations between these important learning processes.

My research contributes to the ITS literature and may have the potential to stimulate important educational data mining work. The insights gained from my experiments could be incorporated into a knowledge tracing model that can help provide adaptive support for learning from MGRs. It would be interesting to investigate whether my theoretical model for learning with MGRs is applicable across domains that use MGRs to support different conceptual interpretations of the same abstract concept. This envisioned knowledge tracing model might thus have broad impact for the development of ITSs across a wide range of domains.

Finally, I contribute to HCI since the Fractions Tutor demonstrates how HCI methods can help the design of an effective system that supports students’ learning in a complex domain.

7 Literature


