Conceptual learning with multiple graphical representations: Intelligent tutoring systems support for sense-making and fluency-building processes

Martina A. Rau

August 2013

CMU-HCII-13-100

Doctoral Dissertation

Human-Computer Interaction Institute

School of Computer Science

Carnegie Mellon University

Pittsburgh, PA USA

Thesis Committee:

Vincent Aleven, Human-Computer Interaction Institute, Carnegie Mellon University (Chair)

Nikol Rummel, Institute of Educational Research, Ruhr-Universität Bochum (Chair)

Ken Koedinger, Human-Computer Interaction Institute, Carnegie Mellon University

Richard Scheines, Department of Philosophy, Carnegie Mellon University

Brian Junker, Department of Statistics, Carnegie Mellon University

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

Most learning environments in the STEM disciplines use multiple graphical representations along with textual descriptions and symbolic representations. Multiple graphical representations are powerful learning tools because they can emphasize complementary aspects of complex learning contents. However, to benefit from multiple graphical representations, students need to engage in a number of learning processes. Educational technologies offer novel opportunities to support these learning processes by making graphical representations interactive and by providing individualized instructional support for students' interactions with them. Yet, these opportunities are under-researched, as most prior research has taken a symbol-systems approach by focusing only on multiple representations that use *different* symbol systems, such as text and one additional graphical representation.

To address the open question of how to enhance students' benefit from *multiple graphical* representations that use *the same* symbol system, I conducted a series of five classroom experiments and lab studies with over 3,000 students in grades 4-6. Each experiment tested the effectiveness of different types of instructional support for students' learning with multiple graphical representations. Experiment 1 compares the effects of multiple over a single graphical representation and the effects of prompting students to self-explain the relation between graphical and symbolic representations. Results show that multiple graphical representations lead to better learning than a single graphical representation, provided that students receive self-explanation prompts. Experiment 2 contrasts sequences of task types and graphical representations. The results show that interleaving task types while blocking graphical representations promotes students' learning from multiple graphical representations more so than interleaving graphical representations while blocking task types. Building on Experiment 2, Experiment 3 investigates whether (in addition to moderately interleaving task types) graphical representations should also be presented in an interleaved, rather than in a blocked fashion. An analysis of learning outcomes and tutor log data demonstrates that interleaving graphical representations (while moderately interleaving task types) enhances students' benefits from multiple graphical representations. Furthermore, Experiment 3 replicates the finding from Experiment 1 that multiple graphical representations lead to better learning than a single one. Experiment 4 investigates the effects of different types of instructional support for connection making between multiple graphical representations. The results show that a combination of support designed to help students actively make

sense of the connections and of support designed to help students become fluent in making these connections is needed for students to benefit from multiple graphical representations, compared to a single graphical representation. Finally, Experiment 5 investigates different sequences of connectional sense-making support and connectional fluency-building support. The results lead to the conclusion that receiving support for making sense of connections first is a prerequisite to students' benefit from subsequent connectional fluency-building support.

A further contribution of my thesis work is the development of an intelligent tutoring system for fractions that leads to significant and robust gains in students' conceptual and procedural knowledge of fractions. In addition to investigating how best to support students' learning with multiple graphical representations, each experiment also served to iteratively improve the Fractions Tutor while employing user-centered techniques. To develop the Fractions Tutor, I made use of a novel methodology to resolve stakeholder conflicts that inevitably arise in complex educational settings.

I consolidate my empirical findings in a novel theoretical framework that describes the learning processes that students perform when learning with multiple graphical representations. This framework extends existing theoretical frameworks, which have solely focused on learning with representations that use different symbol systems (such as text accompanied with one additional graphical representation), rather than on learning with multiple representations using the same symbol system (such as multiple graphical representations). My theoretical framework proposes that in order to benefit from multiple graphical representations, students need to conceptually understand each individual graphical representation and to use each graphical representation fluently to solve domain-specific problems, students need to conceptually understand the connections between different graphical representations, and they need to become fluent in making these connections.

In sum, my thesis work contributes (1) an empirically validated set of instructional design principles for the effective use of multiple graphical representations, (2) a theoretical framework for learning with multiple graphical representations that use *the same* symbol system, (3) an effective tutoring system for fractions learning, and (4) a new methodology for resolving design conflicts that often occur in real educational settings.

Table of Contents

1 Introduction

Instructional materials in science, technology, engineering, and math (STEM) domains universally employ a variety of representations as educational tools: flow diagrams in programming, schemas and tree diagrams in biology, charts and graphs in math – to mention just a few examples. Indeed, a vast amount of literature documents the potential benefit of multiple representations on students' learning (Ainsworth, 1999, 2006; Ainsworth, Bibby, & Wood, 2002; Brenner et al., 1997; de Jong et al., 1998; Eilam & Poyas, 2008; Hwang, Su, Huang, & Dong, 2009; Kordaki, 2010; Solaz-Portolés & Lopez, 2007; van Someren, Boshuizen, & de Jong, 1998). Multiple representations are considered to enhance learning in part because different representations emphasize complementary conceptual aspects of the learning material and have differential effects on mental processing (Cox, 1999; Cromley, Snyder-Hogan, & Luciw-Dubas, 2010; Gagatsis & Elia, 2004; Gegenfurtner, Lehtinen, & Säljö, 2011; Goldman, 2003; Hinze et al., 2013; Kozma, Chin, Russell, & Marx, 2000; Larkin & Simon, 1987; Reed & Ettinger, 1987; Schnotz & Bannert, 2003; Schwartz & Black, 1996; Tabachneck, Leonardo, & Simon, 1997; Zhang, 1997; Zhang & Norman, 1994). Across domains, researchers and instructors recognize the importance of using more than one representation, as "the adherence to […] one visualization – as the ultimate kind of organization for the phenomenon at hand – impedes students' development of […] cognitive flexibility" (Eilam, 2013, p. 69; also see Spiro, Feltovich, Jacobson, & Coulson, 1991).

However, most prior research has taken a *symbol-systems approach* by focusing on learning with text and one additional graphical representation (Ainsworth & Loizou, 2003; Baetge & Seufert, 2010; Bodemer, Plötzner, Bruchmüller, & Häcker, 2005; Butcher & Aleven, 2007; Kuehl, Scheiter, & Gerjets, 2010; Magner, Schwonke, Renkl, Aleven, & Popescu, 2010; Rasch & Schnotz, 2009; Suthers, Vatrapu, Medina, Joseph, & Dwyer, 2008). This focus on learning with text and graphical representations reflects the belief that multiple representations are beneficial because they are presented in *different* symbol systems. Symbol systems are defined by the mental construction processes through with a learner constructs an internal representation of the content that is being depicted (Schnotz & Bannert, 2003, see section 2.1). Textual descriptions are encoded based on their semantic organization, whereas graphical representations are encoded based on their perceptual meaning. The integration of information presented across different symbol systems requires deeper processing of the learning content than when the same content is presented in only one symbol system (e.g., text alone) and thus accounts for the positive effect of multiple representations (Schnotz & Bannert, 2003). By contrast, learning materials across all STEM domains usually include multiple representations of *the same* symbol system: textual descriptions and symbolic notations are typically accompanied by *multiple graphical* representations (Arcavi, 2003; Cook, Wiebe, & Carter, 2007; Kordaki, 2010; Kozma et al., 2000; Urban-Woldron, 2009; Walkington, Nathan, Wolfgram, Alibali, & Srisurichan, 2011). Unfortunately, given the pervasiveness of the symbol-systems approach in the educational psychology literature, prior research fails to address this common scenario found in many educational materials. In the light of prior research, it remains unclear whether multiple graphical representations even *benefit* students' learning. Requiring students to integrate multiple representations that employ the same symbol system (such as multiple graphical representations) might harm students' learning because they lead to cognitive overload in the pictorial part of working memory (Clark & Mayer, 2003; Leikin, Leikin, Waisman, & Shaul, 2013; Mayer, 2003; Mayer, 2005), which might impede learning (Chandler & Sweller, 1991).

It is widely recognized that simply providing learners with multiple representations will not necessarily enhance learning. Research on multiple representations shows that they only enhance learning when accompanied with appropriate instructional support (Baetge & Seufert, 2010; Berthold, Eysink, & Renkl, 2008; Berthold, Faulhaber, Guevara, & Renkl, 2010; Bodemer & Faust, 2006; Bodemer et al., 2005; Butcher, 2006; Gobert et al., 2011; Mayer & Gallini, 1990; Ozcelik, Karakus, Kursun, & Cagiltay, 2009; Plötzner, Bodemer, & Neudert, 2008; Rathmell & Leutzinger, 1991; Schwonke, Berthold, & Renkl, 2009; Schwonke, Ertelt, & Renkl, 2008; Seufert, 2003; Uttal, 2003; van der Meij, 2007; van der Meij & de Jong, 2006). Yet again, given that this research has largely taken a symbol-systems approach, it is an open question how best to best design instructional materials with *multiple graphical* representations, and how to best to provide instructional support for students' robust learning with multiple graphical representations.

Indeed, in reviewing science and math curricula (Bennett, 2004; Corwin, Russell, & Tierney, 1990; Cramer, Behr, Post, & Lesh, 1997a, 1997b; Hake, 2004; Kilpatrick, Swafford, & Findell, 2001; Lappan, Fey, & Fitzgerald, 1998), I found that while they all include a variety of graphical representations, each uses them in different ways. Although many educational standards emphasize the importance of integrating a variety of graphical representations (Halpern et al., 2007; NCTM, 1989, 2000; NETP, 2010; NMAP, 2008; Pashler et al., 2007; Siegler et al., 2010), they do not provide specific guidance as to how to implement them. Therefore, the goal of my dissertation work is to investigate how to use *multiple graphical* representations in a way that most benefits students' learning of robust domain knowledge: knowledge that transfers to novel tasks and lasts over time (Koedinger, Corbett, & Perfetti, 2012).

I investigate this question in the context of an educational technology. Educational technologies offer novel opportunities in supporting students' learning from graphical representations, for instance by making representations interactive (Crawford & Brown, 2003; Durmus & Karakirik, 2006; Gire et al., 2010; Lewalter, 2003; Moyer, Bolyard, & Spikell, 2002; Proctor, Baturo, & Cooper, 2002; Rasch & Schnotz, 2009; Reimer & Moyer, 2005; Suh, Moyer, & Heo, 2005), and by providing individualized support with respect to students' interactions with the representations (Durmus & Karakirik, 2006; Kordaki, 2010; Suh & Moyer, 2007). At the same time, educational technologies increasingly impact education in real classrooms across the United States as internet-based courses are fast expanding (Kim, Kim, & Whang, 2013), as schools are getting more and more access to computers (DeBell et al., 2003) and to the internet (Wells & Lewis, 2006), and as virtual schools become more and more prevalent (Oliver, Osborne, & Brady, 2009). My dissertation research therefore integrates the learning sciences question of how best to support students' robust learning with multiple graphical representations and the educational technology question of how best to capitalize on the novel opportunities that technologies offer to provide instructional support for students' interactions with multiple graphical representations.

Taken together, the contributions of my work fall into four categories. First, my research provides a set of instructional design principles on how to support students' robust learning through principled use of multiple graphical representations (see section 5.1). These instructional design principles are the outcome of a sequence of controlled experiments (see section 4). They address the gap between the common symbol-systems approach in prior educational psychology research and the lack of guidance in common educational materials that include multiple graphical representations that use the same symbol system. As such, the instructional design principles provide guidance for the development of multi-representational educational technologies.

Second, I discuss these principles in the light of a new theoretical framework for learning with multiple graphical representations (see section 2.2 and section 5.1.3). In providing a new theoretical framework, I extend prior frameworks on multiple representations (see sections 2.2 and 5.1). My goal is to explicitly describe the learning processes that facilitate students' learning with multiple representations that use the same symbol system. The theoretical framework builds on existing theoretical frameworks for learning with multiple representations from *different* symbol systems and is the result of a series of experiments (see section 4) that I have conducted on learning with multiple representations of *the same* symbol system. Furthermore, the theoretical framework makes testable predictions some of which I have investigated in my empirical work. Furthermore, the theoretical framework has the potential to stimulate research on learning with multiple graphical representations in a variety of domains.

Third, I developed a successful piece of educational software: an intelligent tutoring system for fractions (see section 5.2). The Fractions Tutor (see section 3) uses multiple graphical representations in a principled way (see section 3.2), leads to robust conceptual learning (see section 3.5), and is usable within the context of real classroom settings (see section 3.3).

Finally, I describe a user-centered design methodology that integrates methods from learning sciences, intelligent tutoring systems, and human-computer interaction (see section 3.3). This methodology served as a basis for the development of the Fractions Tutor. It incorporates the instructional design principles that follow from a sequence of controlled classroom experiments and has been iteratively improved throughout these experiments (see section 4).

These contributions of my research are not to be considered separate; rather, they iteratively build on one another. An overarching theme across these threads of my work is the integration of multiple research perspectives into a coherent program of research. In integrating multiple perspectives, I employ a multi-methods approach that combines quantitative measures of learning outcomes, qualitative assessments of student and teacher satisfaction, a variety of quantitative and qualitative measures of learning processes, such as tutor log data, think-aloud protocols, and interview data, and uses a variety of research methods, including classroom experimentation and data mining. By bringing multiple perspectives together, my work illustrates how the combination of learning sciences and educational technology research yields insights that exceed what can be gained by adhering to either research perspective alone.

The remainder of this thesis is structured as follows. I first introduce the two perspectives on my work, the learning science perspective (see section 1.1), and the educational technology perspective (see section 1.2). Next, I present a new theoretical framework for learning with multiple graphical representations (see section 2), by extending existing frameworks for learning with text and graphical representations. Then, I describe the Fractions Tutor (see section 3): an intelligent tutoring system that is both the platform and the outcome of my empirical research. I then describe a series of classroom experiments that I conducted with the Fractions Tutor to investigate learning with multiple graphical representations (see section 4). I focus in detail on a final lab-based experiment in this series, which investigated how sense-making processes and fluency-building processes in connection making between multiple graphical representations interact (see section 4.5). Finally, I draw conclusions from these experimental studies in relation to the learning sciences perspective (see section 5.1) and the educational technology perspective of my work (see section 5.2). After discussing the limitations of my research and perspectives for future work (see section 5.3), I end by discussing the merit of integrating complementary perspectives by using a multi-methods approach to empirical research (see section 5.4).

1.1 Learning sciences perspective

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; Hinze et al., 2013; Kozma et al., 2000; Löhner, van Joolingen, & Savelsbergh, 2003; Rasch & Schnotz, 2009; Schnotz & Bannert, 2003) and enhance different kinds of cognitive processes and strategies (Ainsworth, 2006; Cromley et al., 2010; Gagatsis & Elia, 2004; Larkin & Simon, 1987; Lewalter, 2003; Nistal, Van Dooren, Clarebout, Elen, & Verschaffel, 2010; Reed & Ettinger, 1987; Schnotz & Bannert, 2003; Schwartz & Black, 1996; Zhang, 1997; Zhang & Norman, 1994), instructional materials tend to use not just a single representation, but multiple. Indeed, in the educational psychology literature, there is extensive experimental evidence that multiple representations can lead to better learning outcomes than a single representation (Ainsworth et al., 2002; Brenner et al., 1997; Eilam & Poyas, 2008; Schnotz & Bannert, 2003).

By and large, the educational psychology literature on learning with multiple representations has taken a *symbol-systems approach*: research has mostly focused on learning with textual descriptions and one additional graphical representation (Ainsworth & Loizou, 2003; Baetge & Seufert, 2010; Bodemer et al., 2005; Butcher & Aleven, 2008; Butcher, 2006; Eilam & Poyas, 2008; Eitel, Scheiter, & Schüler, 2013; Kuehl et al., 2010; Magner et al., 2010; Mason, Pluchino, & Tornatora, 2013; Rasch & Schnotz, 2009; Suthers et al., 2008). Textual descriptions and graphical representations differ in the symbol system they use (Schnotz & Bannert, 2003). Text is organized verbally and is interpreted via semantic processing of the text-based structure, leading to a propositional internal representation of the content. Graphical representations are organized according to visual structures and are interpreted via analog perception and thematic selection, leading to an analog internal representation. Under the symbol-systems approach, the advantage of multiple representations has been attributed to the fact that they stimulate deeper processing by requiring learners to integrate information across these different symbol systems (i.e., by integrating propositional representations and analog internal representations). However, instructional materials found in real educational settings are typically more complex. Learning materials across a wide range of domains contain *multiple graphical* representations in addition to text and symbolic representations (van Someren et al., 1998), for instance in math (Arcavi, 2003; Cheng, 1999; Noss, Healy, & Hoyles, 1997; Pape & Tchoshanov, 2001; Rathmell & Leutzinger, 1991), chemistry (Kozma et al., 2000; Kozma & Russell, 2005; Stieff, Hegarty, & Deslongchamps, 2011; Zhang & Linn, 2011), biology (Cook et al., 2007; Simons & Keil, 1995), physics (Larkin & Simon, 1987; Lewalter, 2003; Urban-Woldron, 2009; van der Meij & de Jong, 2006), engineering (Nathan, Walkington, Srisurichan, & Alibali, 2011; Walkington et al., 2011), and programming (Baetge & Seufert, 2010; Kordaki, 2010). Thus, all of these domains typically incorporate multiple representations that use the same symbol system. Whether prior research on learning with multiple representations from *different* symbol systems generalizes to learning with multiple representations from *the same* symbol system remains an open question.

In spite of the well-documented promise of learning with multiple representations, research has not always succeeded in demonstrating their advantage on students' learning (Baetge & Seufert, 2010; Kuehl et al., 2010; Mayer & Gallini, 1990; Schnotz & Bannert, 2003; Schoor & Bannert, 2010). Students' benefits from multiple representations rely on their ability to understand each individual representation (Ainsworth, 2006; Eilam, 2013), and their ability to make connections between them (Ainsworth, 2006; de Jong et al., 1998; Gobert et al., 2011; Gutwill, Frederiksen, & White, 1999; Özgün-Koca, 2008; Rathmell & Leutzinger, 1991; Superfine, Canty, & Marshall, 2009; Uttal, 2003; van der Meij, 2007). Unfortunately, we do not yet fully understand whether students need to engage in the same learning processes when learning with multiple representations using *different* symbol systems as when learning with multiple representations using *the same* symbol system. Consequently, we do not know how best to support students in learning with multiple graphical representations, although it seems reasonable to believe that such support is integral to their benefit from them. Therefore, research is needed that focuses on the common case of learning with multiple graphical representations in order to 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 et al., 2012).

Fractions instruction is one of the many domains in which multiple graphical representations, such as circles, rectangles, and number lines are used extensively (NMAP, 2008; Siegler et al., 2010). Different graphical representations 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; Cramer, 2001; Cramer & Henry, 2002; Cramer & Wyberg, 2009; Cramer, Wyberg, & Leavitt, 2008; Cramer, Post, & delMas, 2002; Lamon, 1999; Reimer & Moyer, 2005). 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 (Cramer & Wyberg, 2009; Cramer et al., 2002; Lamon, 1999). Linear models (e.g., number lines) depict fractions in the context of measurement and demonstrate that fractions can lie between any two whole numbers (Lamon, 1999; Siegler, Thompson, & Schneider, 2011). By contrast, linear models do not have an inherent unit. Rather, the unit is defined by convention: the length between 0 and 1 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 multiple graphical representations is to help students understand the complex topic of fractions by highlighting these complementary conceptual aspects (NMAP, 2008; Pashler et al., 2007; Siegler et al., 2010). This practice makes fractions a suitable domain for research on learning with multiple graphical representations: as in many other STEM domains, various graphical representations, each with a different conceptual focus, are used in conjunction with textual descriptions and symbolic representations to enhance learning of rather abstract concepts.

My research addresses these open questions about how best to support students' robust learning with multiple graphical representations. Thereby, my research extends prior research on learning with multiple representations using *different* symbol systems to learning with *multiple graphical* representations using *the same* symbol system, which is – as just argued – a common scenario in real-world instructional materials. Given that processes involved in learning with representations from different symbol systems (i.e., when integrating propositional representations and analog internal representations in the case of textual descriptions and graphical representations) may differ from the processes involved in learning with multiple graphical representations (i.e., when integrating multiple analog internal representations), my research provides fundamentally novel insights that may have an impact on educational materials across a large range of domains.

To achieve these goals, I conducted a sequence of experiments on learning with multiple graphical representations of fractions (Rau, Aleven, & Rummel, 2009, 2010; Rau, Aleven, & Rummel, 2013b; Rau, Aleven, Rummel, & Rohrbach, 2012, 2013; Rau & Pardos, 2012; Rau, Rummel, Aleven, Pacilio, & Tunc-Pekkan, 2012; see section 4). The goal of each experiment is (1) to investigate how best to implement multiple graphical representations and instructional support for learning with them so as to enhance students' learning of robust domain knowledge, and (2) to develop and iteratively revise the Fractions Tutor, a piece of educational software for fractions learning, which incorporates these instructional design principles. The experiments served as a basis for both the development and the refinement of a theoretical framework for learning with multiple graphical representations that extends existing theoretical frameworks on learning with multiple representations that have taken a symbol-systems approach (see section 2). Both the theoretical framework and the experimental studies provide the foundation for the design of educational software for fractions learning (see section 3).

1.2 Educational technology perspective

Educational technologies provide novel opportunities to support students' learning with multiple graphical representations. A variety of studies document the potential advantages of allowing students to work with dynamic, interactive representations in mechanics (Bodemer, Plötzner, Feuerlein, & Spada, 2004; Plötzner et al., 2008), chemistry (Chiu & Linn, 2012; Zhang & Linn, 2011), physics (Gire et al., 2010; Hwang et al., 2009; Lewalter, 2003; van der Meij, 2007), algebra (Suh & Moyer, 2007), and statistics (Plötzner et al., 2008). In math education, virtual manipulatives have recently gained attention (Crawford & Brown, 2003; Durmus & Karakirik, 2006; Moyer et al., 2002): virtual manipulatives are dynamic, interactive graphical representations embedded in educational technologies that students can manipulate in various ways. Several studies in the domain of fractions argue that virtual manipulatives can enhance students' learning (Lamberty & Kolodner, 2002; Proctor et al., 2002; Reimer & Moyer, 2005; Suh et al., 2005), and that they are at least as effective as physical manipulatives traditionally used for fractions instruction in classrooms (Roussou, Oliver, & Slater, 2006).

Interactive graphical representations can provide individualized support for students' interactions with graphical representations designed to help them acquire cognitive competencies that are prerequisite for their benefit from multiple graphical representations. Yet, these opportunities are under-researched, leaving developers of educational technologies with little guidance on how best to implement instructional support for learning with multiple graphical representations. By investigating learning with multiple graphical representations in the context of educational technologies, my research addresses these shortcomings.

For this purpose, I developed software that uses multiple, interactive graphical representations of fractions. In doing so, I made use of a particularly successful educational technology: a type of Cognitive Tutor (Koedinger & Corbett, 2006; Ritter, Anderson, Koedinger, & Corbett, 2007). Cognitive Tutors are grounded in cognitive theory and artificial intelligence. They pose rich problem-solving tasks to students and provide individualized support at any point during the problem-solving process. At the heart of Cognitive Tutors lies a cognitive model of students' problem-solving steps. The model serves as a basis for individualized support given to students throughout the learning process (Corbett & Anderson, 2001; Corbett & Trask, 2000; Koedinger & Corbett, 2006). Cognitive Tutors have been shown to lead to significant learning gains in a variety of studies (Corbett & Anderson, 2001; Corbett & Trask, 2000; Corbett, Koedinger, & Hadley, 2001; Koedinger & Corbett, 2006; VanLehn, 2011), and are currently being used in close to 3,000 U.S. schools. The Fractions Tutor is a specific type of Cognitive Tutor: an example-tracing tutor (Aleven, McLaren, Sewall, & Koedinger, 2009a) that provides step-by-step guidance in the form of feedback and on-demand hints in the same way as other Cognitive Tutors do, but instead of relying on a rule-based cognitive model, it does so based on generalized examples of correct and incorrect solution paths rather than on a rule-based cognitive model of student behavior.

In developing the Fractions Tutor, my goal was to develop software that is usable within the constraints of real educational contexts and that addresses the goals and needs of multiple stakeholders in these contexts (e.g., students, teachers, superintendents). Different stakeholders often have different goals. Furthermore, there are typically significant resource limitations, so that design goals (even if they were agreed upon by all stakeholders) need to be traded off against each other. Unfortunately, existing design methodologies for educational technologies (Bereiter & Scardamalia, 2003; Design-based Research Collective, 2003; Jackson, Krajcik, & Soloway, 1998; Koedinger, 2002; Soloway et al., 1996; van Merriënboer, Clark, & de Croock, 2002) do not provide guidance on how to resolve such design conflicts. This problem may in part be attributed to the fact that many different perspectives are relevant to the development of educational technologies. While each methodology mainly focuses on one of these perspectives, they rarely integrate these different perspectives. In particular, some methodologies focus on user-centered design (Design-based Research Collective, 2003; Jackson et al., 1998; Soloway et al., 1996), others incorporate learning sciences (Bereiter & Scardamalia, 2003) and cognitive psychology and cognitive science research (Koedinger, 2002; Mayer, 2003; van Merriënboer et al., 2002). Since these different types of methodologies rarely reference one another, developers often have to rely on ad-hoc methods to resolve conflicts that inevitably arise in the interdisciplinary field of educational technology. For instance, a math teacher who wants to help students learn deeply may provide complex real-world problems (Bereiter & Scardamalia, 2003). Yet, van Merriënboer and colleagues (2002) suggest practicing part-tasks: discrete tasks that are necessary for the completion of complex problems (e.g., practicing math facts). At the same time, students find complex problems interesting, but the teacher might worry that their learning is jeopardized because the problems do not provide just-in-time feedback (van Merriënboer et al., 2002).

Paying attention to such conflicting goals is a crucial prerequisite to developing effective educational software. If we fail to address stakeholders' competing goals, students may dislike the software because it is either boring or too challenging, or teachers – who might well believe it will help their students learn deeply – fail to use the software within the constraints of their day-to-day job, which requires them to prepare students for standardized tests and manage a classroom. However, if we succeed in integrating stakeholders' needs within the constraints of their contexts into the design of educational software, dissemination and long-term success of the software, and by consequence, students' learning outcomes, will hugely benefit. What is needed is a principled methodology that developers can apply to resolve such conflicts.

Cognitive Tutors are particularly suitable for developing a methodology to resolving design conflicts as their development follows a well-described design process that integrates design recommendations originating from a number of fields, including human-computer interaction, learning science, and education research (Corbett, Koedinger, & Anderson, 1997). I extend this process by providing a new approach for resolving conflicting design recommendations and constraints. In particular, my methodology combines focus groups and affinity diagramming to develop a goal hierarchy, parametric experiments, and cross-iteration studies. The novelty of this methodology lies in a principled combination of methods that originate in a variety of disciplines, including from human-computer interaction, learning sciences, and intelligent tutoring systems research.

Taken together, the success of the Fractions Tutor, like that of other Cognitive Tutors (Corbett, Kauffman, Maclaren, Wagner, & Jones, 2010; Ogan, Aleven, & Jones, 2008; Ritter et al., 2007), has been shaped by incorporating stakeholder goals into the design process. In doing so, I employed a principle-based methodology, which combines learning sciences and intelligent tutoring systems research with a user-based design perspective. I believe that this methodology is not unique to the domain of fractions or Cognitive Tutors in particular, but that can inform the development of a wide range of educational software.

1.3 Summary

In summary, my research integrates multiple perspectives that complement each other. First, my research provides a set of instructional design principles on how to employ multiple graphical representations most effectively so that they enhance students' robust learning of domain knowledge. To this end, I conducted several classroom experiments and lab studies with over 3,000 students. Each served to investigate theoretically motivated questions about how best to support students' learning through the use of multiple graphical representations. From this perspective, my goal in developing the Fractions Tutor was to use it as a research platform to study learning sciences questions about learning with multiple graphical representations. Second, I developed a novel theoretical framework for learning with multiple graphical representations in conjunction with the experimental studies. The theoretical framework extends prior theoretical perspectives on learning with multiple representations that use *different* symbol systems to learning with multiple graphical representations that use *the same* symbol system. I discuss the findings from my experimental studies from the perspective of the theoretical framework while highlighting new research questions for future research that the theoretical framework raises. Third, my goal was to develop a successful intelligent tutoring system that addresses an educational problem and that is usable within real educational contexts in which conflicts between stakeholder goals inevitably arise. From this perspective, the Fractions Tutor can be viewed as the outcome of my research. Each experimental study also served to iteratively improve the Fractions Tutor. Finally, the in developing the Fractions Tutor, I developed a methodology to address conflicts between competing stakeholder goals that may arise in complex educational settings. My work illustrates how the use of a multi-methods approach that combines learning outcome measures with process-level measures, can use one perspective to enhance the benefit of the other perspective: the different perspectives of my research do not stand side by side but complement each other.

Altogether, in combining learning sciences, intelligent tutoring systems, and user-centered design perspectives, my research results in (1) a set of instructional design principles to guide the development of a range of multi-representational educational technologies, (2) an empirically motivated theoretical framework for learning with multiple graphical representations that use the same symbol system, (3) an effective tutoring system for fractions learning, and (4) a new methodology for resolving design conflicts that often occur in real educational settings.

2 Theoretical Framework

To develop educational software that supports learning through adequate use of multiple graphical representations, it is crucial to first reflect on the cognitive processes involved in learning with multiple graphical representations. Simply integrating multiple graphical representations into educational software is unlikely to lead to optimal learning gains (Ainsworth, 2006; de Jong et al., 1998; Kim et al., 2013). Rather, their success in supporting robust learning stands and falls with appropriate instructional support to help students understand individual representations (Ainsworth, 2006; Eilam, 2013), and the connections between them (Ainsworth, 2006; de Jong et al., 1998; Gobert et al., 2011; Gutwill et al., 1999; Özgün-Koca, 2008; Rathmell & Leutzinger, 1991; Superfine et al., 2009; Uttal, 2003; van der Meij, 2007).

There are a number of theoretical frameworks that describe the cognitive processes involved in learning with multiple representations, and that provide guidance for the design of multi-representational learning materials. Yet, these frameworks take a *symbol-systems approach*: they are based on research on learning with multiple representations using *different* symbol systems, such as text and one additional graphical representation. None of them specifically focuses on learning with *multiple graphical* representations using *the same* symbol system. In this section, I first describe a number of existing frameworks for learning with multiple representations, while explicitly pointing out their shortcomings and why their predictions may not explain the hypothesized advantage of multiple graphical representations over a single graphical representation. I then provide a new theoretical framework for learning with *multiple graphical* representations that addresses these shortcomings and provides guidance on how to design instructional support for students' learning with multiple graphical representations. I developed and further refined this theoretical framework along with empirical work. Even though here, I present the theoretical framework before describing my experimental studies, the framework is to be considered an outcome of my empirical work, as well as the motivation for some of my experimental studies described in section 4.

2.1 Existing theoretical frameworks for learning with multiple representations

When discussing theoretical frameworks 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" (Gilbert, 2008; Zhang, 1997, p. 180). Internal representations, on the other hand, are knowledge structures in memory, such as schemas or production rules (Gilbert, 2008; Zhang, 1997). As learners understand external representations, they form internal representations, which they then, in turn, integrate into a mental model (Gilbert, 2008; Zhang, 1997; Zhang & Norman, 1994). Since different representations have complementary strengths (Ainsworth, 2006; Cromley et al., 2010; Gagatsis & Elia, 2004; Hinze et al., 2013; Kozma et al., 2000; Larkin & Simon, 1987; Lewalter, 2003; Löhner et al., 2003; Nistal, Van Dooren, Clarebout, Elen, & Verschaffel, 2010; Rasch & Schnotz, 2009; Reed & Ettinger, 1987; Schnotz & Bannert, 2003; Schwartz & Black, 1996; Zhang, 1997; Zhang & Norman, 1994), the effectiveness of multiple external representations lies in their potential to help students form more accurate mental models of the domain.

Building on the distinction between internal and external representations, Schnotz and Bannert's (2003) theoretical framework (Fig. 1) proposes that text and graphical representation lead to different types of internal representations – by virtue of using different symbol systems. Text uses a verbally organized symbol system and is processed semantically through an analysis of its semantic structure, leading to a propositional internal representation. Graphical representations, on the other hand, use a pictorial symbol system and are processed perceptually, leading to an analog internal representation. During mental model formation, learners integrate both types internal representations via structure mapping (Gentner & Markman, 1997). Integrating information from different symbol systems requires deep conceptual processing, as students have to map semantic propositions to corresponding parts of a pictorial, analog internal representation. According to Schnotz and Bannert (2003), the integration across different symbol systems explains why learning with text and graphical representation leads to better learning outcomes than learning with text alone.

In line with the Dual Channel theory (Paivio, 1986), and building on Schnotz and Bannert's framework of text and graphic comprehension (Schnotz & Bannert, 2003), Mayer and Moreno's

Cognitive Theory of Multimedia Learning (Mayer, 2003; Mayer, 2005; Mayer & Moreno, 1998, 2003) assumes that verbal and pictorial information are processed in different information channels (Fig. 2). Even though text is often presented visually (i.e., in written form), it is encoded into a verbal model within working memory (see Fig. 2). Since the capacity of each part of working memory is limited but the capacity of both together is additive (Chandler & Sweller, 1991), learning with both text and graphical representation makes better use of the learner's working memory capacity than learning with either type of representation alone. Furthermore, active integration of the verbal model and the pictorial model into one coherent mental model requires deeper conceptual processing of the content, which leads to better learning, compared to learning with a single representation.

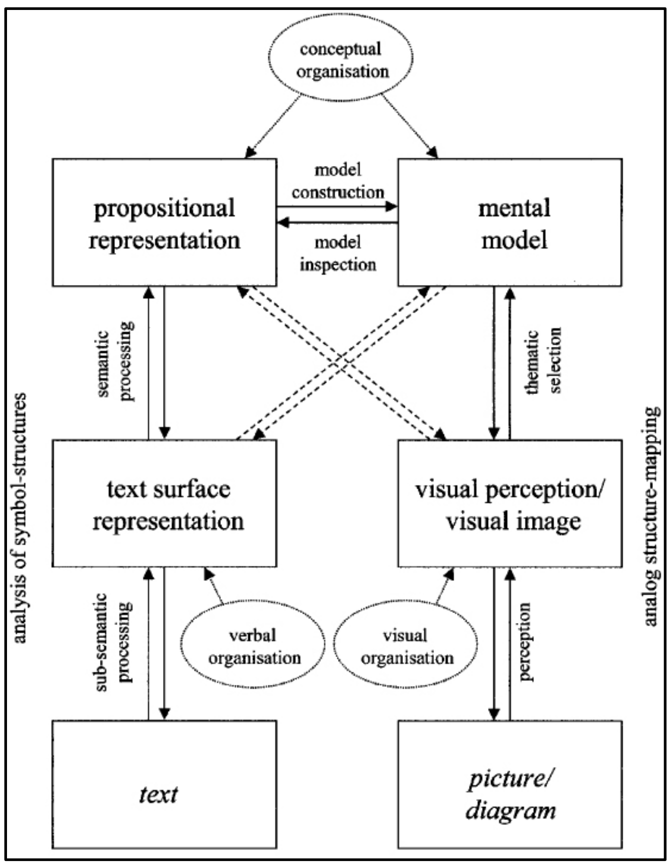


Fig. 1. Mental model integration in Schnotz and Bannert's (2003) theoretical framework of text and graphic comprehension.

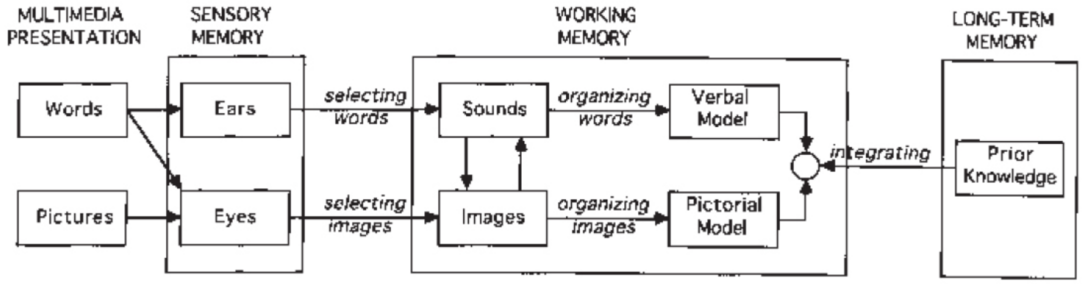


Fig. 2. Dual channel processing of verbal and pictorial information in Mayer and Moreno's Cognitive Theory of Multimedia Learning.

While the theoretical framework by Schnotz and Bannert (2003) and the Cognitive Theory of Multimedia Learning (Mayer, 2003; Mayer, 2005; Mayer & Moreno, 2003) can explain why representations using *different* symbol systems (i.e., text and one graphical representation) lead to better learning than either alone, they do not predict an advantage of *multiple graphical* representations compared to a single graphical representation. Different graphical representations use *the same* symbol system and consequently initiate the same encoding processes as learners form in internal (pictorial, analog) representation of the external (graphical) representations. Schnotz and Bannert (2003) would therefore not predict an advantage of multiple graphical representations provided in addition to text (when compared to a single graphical representation provided in addition to text): multiple graphical representations do not involve additional symbol systems from which learners have to integrate information. Furthermore, since all graphical representations are encoded into a pictorial model in working memory, multiple graphical representations do not increase a learner's cognitive capacity. The Cognitive Theory of Multimedia Learning might even predict cognitive overload in the pictorial part of working memory if multiple graphical representations are provided in addition to text (as opposed to only a single graphical representation in addition to text), which might hamper learning.

Ainsworth's (2006) Design-Functions-Tasks framework describes cognitive competences involved in learning with multiple *external* representations. Although the Design-Functions-Tasks framework does not explicitly focus on learning with *graphical* representations specifically, it generalizes to learning with multiple graphical representations. 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 some common structure: both number line and circle depict the knowledge components 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 of a circle showing 1/2, and a number line showing 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 length of the 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. The Design-Functions-Tasks framework 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.

Although the Design-Functions-Tasks framework can be extended to explain advantages of multiple graphical representations over a single graphical representation, its empirical basis is research that has been conducted under the symbol-systems approach. Hence, it does not explicitly specify the learning processes that enhance students' benefit from multiple graphical representations in particular, or how to provide instructional support for them. Most research on learning with multiple representations, for instance, concludes that students need to make connections between representations of *different* symbols systems, such as text and graphical representation (Bodemer & Faust, 2006; Bodemer et al., 2004; Plötzner, Bodemer, & Feuerlein, 2001; Plötzner et al., 2008; Schwonke et al., 2008; Seufert, 2003), or between symbolic and graphical representation (van der Meij & de Jong, 2006). It remains an important open question which learning processes play a role in learning with multiple representations of *the same* symbol system, that is, with multiple graphical representations, as they are commonly used in learning materials across a variety of domains.

Building on the prior work reviewed in this section, I provide a new theoretical framework that describes the learning processes involved in learning with multiple graphical representations.

2.2 A new theoretical framework for learning with multiple graphical representations

In describing my theoretical framework, I adopt Anderson's (1993) view that frameworks make general claims that cannot be tested empirically without making additional assumptions. I consider the purpose of the theoretical framework to be the stimulation of further research across STEM domains that use multiple graphical representations. The framework describes learning processes that (arguably) enhance robust learning of domain knowledge with multiple graphical representations. To deduct testable predictions for a given domain, one needs to specify these learning processes by describing corresponding learning events for the given domain. I illustrate how one might derive testable predictions from this framework for the domain of fractions below, hoping to inspire further research in this area.

2.2.1 Processes involved in learning with multiple graphical representations

In extending previous work on learning with multiple external representations, I draw on Koedinger and colleagues' (2012) Knowledge-Learning and Instruction (KLI) framework, which describes processes involved in learning of robust domain knowledge. In most domains, learning involves *sense-making processes* (among memory and fluency-building processes, and induction and refinement processes; Koedinger et al., 2012). In applying Koedinger and colleague's description of sense-making processes to the case of multiple graphical representations, I define sense-making processes as learning processes that lead to principled understanding of connections between multiple graphical representations based on their knowledge components. Knowledge components are "acquired units of cognitive function or structure that can be inferred from performance on a set of related tasks" (Koedinger et al., 2012, p. 764).

I propose that two types of sense-making processes are involved in learning with multiple graphical representations. In order to effectively acquire domain knowledge with multiple graphical representations, students need to engage in sense-making processes to develop conceptual understanding of each individual graphical representation (henceforth *representational sense-making processes*, leading to *representational understanding*) and sense-making processes to develop conceptual understanding of the connections between multiple graphical representations (henceforth *connectional sense-making processes* leading to *connectional understanding*).

Engaging in *representational sense-making processes* means to relate the knowledge components involved in each graphical representation (e.g., the number of colored sections) to the abstract concept they represent (e.g., the numerator) via structure mapping (Gentner & Markman, 1997). That is, students need to relate each graphical representation to the domain-specific knowledge components it depicts (e.g., Ainsworth, 2006; Noss et al., 1997). Understanding graphical representations has been shown to be a difficult task in many domains, such as fractions (Lamon, 1999), statistics (Baker, Corbett, & Koedinger, 2002), algebra (Friel, Curcio, & Bright, 2001; Kaput, 1989; Preece, 1993), chemistry (Kozma & Russell, 2005), and biology (Eilam, 2013). Understanding graphical representations is generally recognized as an important educational goal in the domain of fractions (Siegler et al., 2010; NCTM, 2010), geometry and algebra (NMAP, 2008), and other math domains (Pape & Tchoshanov, 2001). This notion of understanding individual graphical representations includes understanding its format (Ainsworth, 2006; Eilam, 2013), understanding the operators a graphical representation uses (Ainsworth, 2006), understanding the relation between the graphical representation and the domain (Ainsworth, 2006; Eilam, 2013; Kozma & Russell, 2005), and the ability to use the graphical representation to solve a task in the domain (de Jong et al., 1998; Nistal, Van Dooren, Clarebout, Elen, & Verschaffel, 2009).

Engaging in *connectional sense-making processes* means to establish relations between corresponding knowledge components of different graphical representations. The ability to make connections between multiple representations is 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). For example, Bodemer and Faust (2006) show that students who receive support to interactively relating text-based descriptions of heat pumps with corresponding parts in a diagrammatic representation show higher learning gains than students who do not receive such support.

However, learning does not only rely on conceptual 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 (Koedinger et al., 2012). Often, fluency is considered as the ability to retrieve facts from memory (Arroyo, Royer, & Woolf, 2011). In my thesis work, on the other hand, I focus on perceptual fluency, which has been described as the ability to “extract information [...] as the result of experience and practice” (Gibson, 1986, p. 3). Kellman and colleagues (2009) describe perceptual fluency as the ability to fast and effortlessly pick up relevant features and structural relations that define important classifications” (p. 55), and as the ability to “[extract] information more quickly and automatically with practice” (Kellman, Massey, & Son, 2009, p. 3). This type of fluency is an important aspect of learning of domain knowledge (Kellman et al., 2008; Kellman et al., 2009), it happens via unconscious forms of learning (Fahle & Poggio, 2002), and is to be distinguished from conceptual and procedural learning (Kellman & Garrigan, 2009). Fluency-building processes result from experience with the perceptual properties of graphical representations and lead to readily accessible perceptual knowledge about individual graphical representations and about the connections between multiple graphical representations.

I propose that two types of fluency-building processes are involved in learning with multiple graphical representations: processes that lead to fluency in using individual graphical representations (henceforth *representational fluency-building processes* leading to *representational fluency*), and processes that lead to fluency in making connections between different graphical representations (henceforth *connectional fluency-building processes* leading to *connectional fluency*).

Representational fluency describes the ability to quickly and effortlessly identify the information a given graphical representation shows and to use it to solve domain-specific tasks (Kellman et al., 2008; Kellman et al., 2009). A student who has *connectional fluency* can quickly and effortlessly relate different graphical representations by judging “at a glance” that two graphical representations show the same fraction (Kellman et al., 2008; Kellman et al., 2009), rather than reasoning about their constituent knowledge components (i.e., based on reasoning about the numerators and the denominators of the fraction shown in each graphical representation being the same). Perceptually fluent learners can treat one graphical representation as a single perceptual chunk, which allows them to perform domain-relevant tasks quickly and effortlessly with multiple graphical representations.

2.2.2 Deriving domain-specific predictions from the theoretical framework

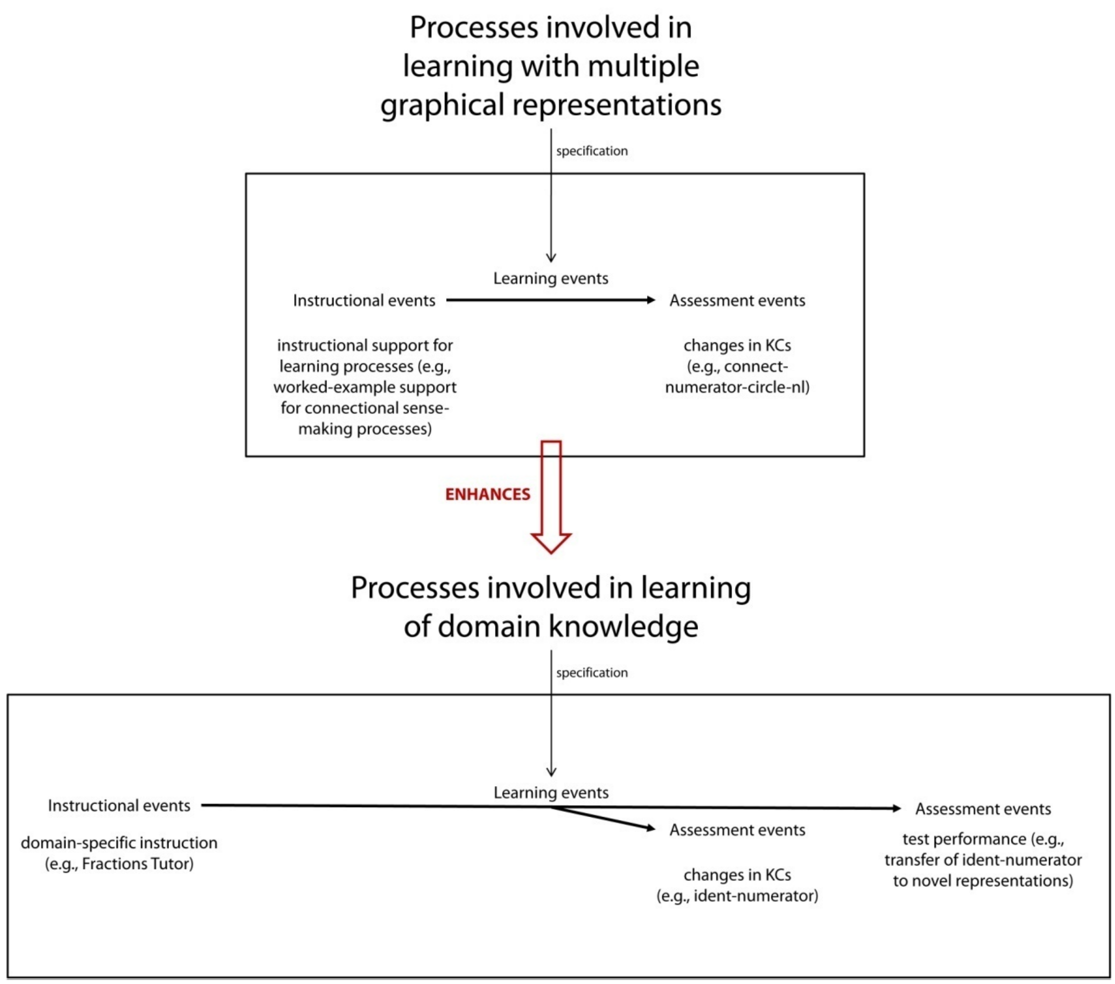


Fig. 3. Specification of processes involved in learning with multiple graphical representations and in learning of domain knowledge for the domain of fractions.

To derive testable predictions from this theoretical framework, I distinguish between two levels of learning: processes involved in learning of *domain knowledge*, and processes involved in learning with *multiple graphical* representations. The central claim of my theoretical framework is that instructional interventions designed to promote processes involved in learning with multiple graphical representations enhance students' benefit from instruction designed to promote learning of the domain knowledge, as illustrated in Fig. 3. Domain knowledge is not specific to the particular graphical representations, but is more general – it includes knowledge that can be used to work with graphical representations, and knowledge that can be used without graphical representations. To make testable predictions for a given domain, one first needs to specify both types of learning processes that students are expected to engage in, by specifying corresponding learning events, as described in Koedinger and colleague's (2012) KLI framework. According to KLI, learning events are changes in knowledge components that happen in response to instructional events and can be inferred from assessment events. To make specific predictions for processes involved in learning of *domain knowledge*, one needs to specify learning events that correspond to a given learning process for the domain of interest. In the domain of fractions, a student should learn how to identify the numerator of a fraction given a graphical representation. (We might call the corresponding knowledge component "ident-numerator). This learning event is supported by instructional events designed to help students acquire this knowledge component (e.g., by having students work on steps in tutor problems that ask them to identify numerators shown in graphical representations). Whether or not the learning event took place can be inferred from assessment events designed to measure whether the student has acquired the knowledge component "ident-numerator" (i.e., the ability to identify the numerator of a fraction given a graphical representations). If the student's knowledge is robust, we expect that this knowledge transfers to novel tasks, for example when the student is required to identify the numerator of a novel graphical representation he/she has never encountered before.

The novelty of my theoretical framework lies in the relation between the processes involved in learning with *multiple graphical* representations described above (see section 2.2.1) to the processes involved in learning of domain knowledge. Let us consider one of the proposed processes involved in learning with multiple graphical representations: connectional sense-making processes. In the domain of fractions, one example of connectional sense-making processes is that a student conceptually understands that the numerator shown by a circle as the number of shaded pieces corresponds to the number of sections between 0 and the dot in the number line. This specific sense-making process constitutes the learning event, since the student acquires the knowledge component "connect-numerator-circle-nl" – a learning event that can be inferred from observable assessment events (e.g., the student correctly maps numerators between circles and number lines). The hypothesis derived from my theoretical framework is that this learning events enhances students' learning of the corresponding domain-specific knowledge component, "ident-numerator". By abstracting across multiple instantiations of the knowledge component "ident-numerator" (e.g., for both circles and number lines), the student becomes more flexible in applying the knowledge component "ident-numerator" to different types of graphical representations – as opposed to acquiring a specific knowledge component "ident-numerator-circle" that only applies to circle representations. Having acquired the flexible knowledge component "ident-numerator" allows the student to transfer this knowledge to novel tasks, for instance to identify the numerator in a set representation, without ever having received instruction about how to identify the numerator depicted in a set.

More broadly speaking, instructional events designed to support learning events that correspond to processes involved in learning with multiple graphical representations (as described in section 2.2.1) should enhance students' learning of domain knowledge. For instance, instructional events designed to support connectional sense-making processes (e.g., worked-example support, see Experiment 4, section 4.4) should enhance students' benefit from instructional events designed to support domain learning. Specifically, students should show better learning outcomes if they learned with a version of the Fractions Tutor (i.e., an intervention designed to support domain learning) that included instructional support for connectional sense-making processes than if they learned with a version of the Fractions Tutor that does not include such support.

2.2.3 Processes involved in learning with multiple graphical representations of fractions

Fig. 4 illustrates the learning processes described above (see section 2.2.1) using three simplified examples of fractions knowledge components involving the concepts of numerator ("ident-numerator"), denominator ("ident-denominator"), and the unit of a fraction (i.e., what the fraction is taken of; "ident-unit").

The left part of Fig. 4 illustrates these *representational sense-making processes* in developing *representational understanding* using the task of identifying a fraction depicted in graphical representations as an example. Here, students are presented with external graphical representations (e.g., a circle, a rectangle, and a number line). To understand how a graphical representation depicts fractions, for instance, students need to learn what component of each graphical representation corresponds to the knowledge components numerator, denominator, and unit of the fraction (Charalambous & Pitta-Pantazi, 2007; Cramer, 2001; Kieren, 1993; Lamon, 1999). In the circle and rectangle, the numerator corresponds to the number of equally sized colored sections ("ident-numerator-circle" and "ident-numerator-rectangle", the denominator to the total number of sections (“ident-denominator-circle” and “ident-denominator-rectangle”), and the unit to the shape of what the fraction is defined in relation to (e.g., a whole circle or a 1-inch long rectangle; “ident-unit-circle” and “ident-unit-rectangle”). For the number line, the numerator is the number of sections between 0 and the dot (“ident-numerator-nl”), the denominator the number of sections between 0 and 1 (“ident-denominator-nl”), and the unit is always the length between 0 and 1 (“ident-unit-nl”). During this sense-making process, students form an internal representation of the knowledge components depicted in each graphical representation. For example, the internal representation of a circle includes the knowledge components “ident-numerator-circle”, “ident-denominator-circle”, and “ident-unit-circle”.

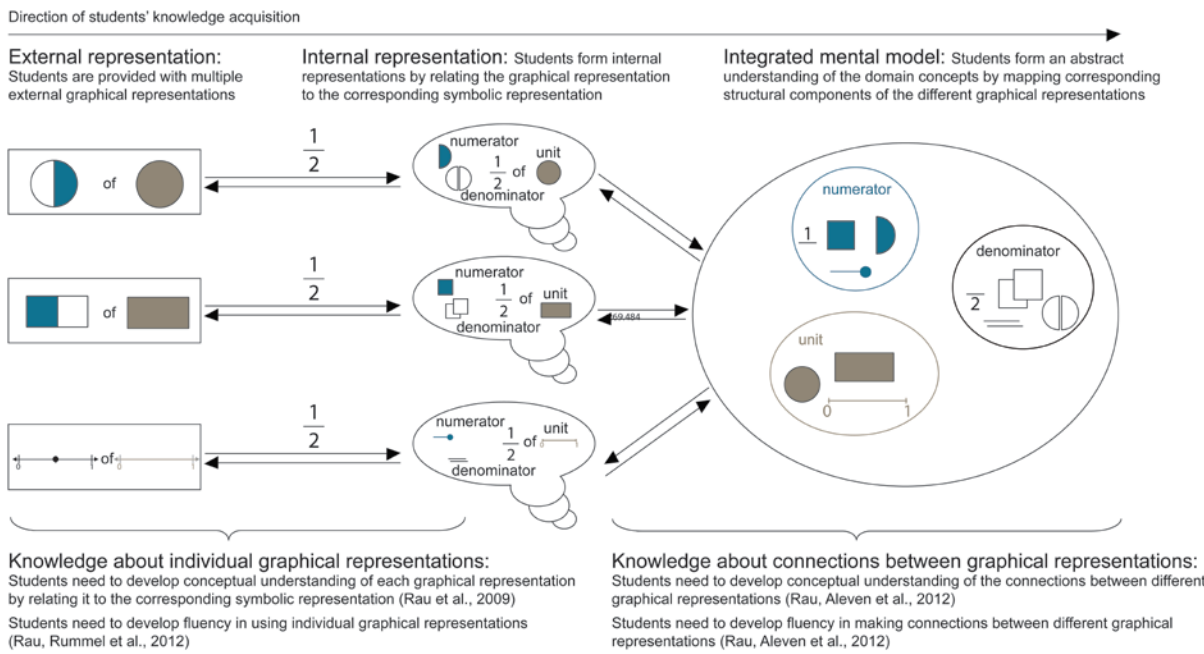


Fig. 4. Theoretical framework for sense-making processes and fluency-building processes in developing knowledge about individual graphical representations and knowledge about connections between multiple graphical representations.

The right part of Fig. 4 illustrates *connectional sense-making processes*. A student might reason that a circle with one colored section shows the numerator of 1 (“ident-numerator-circle”), and that a number line with one section between 0 and the dot shows a numerator of 1 (“ident-numerator-nl”; that a circle with two total sections shows a denominator of 2 (“ident-denominator-circle”), that the number line with two sections between 0 and 1 shows a denominator of 2 (“ident-denominator-nl”). Moreover, the student acquires a mapping of the corresponding components between the numerator in a circle and in a number line (“connect-numerator-circle-nl”) and between the denominator in a circle and in a number line (“connect-denominator-circle-nl”). By making these inferences, the student acquires new and more flexible knowledge components that apply to both circles and number lines and that enables him/her to reason that, since both graphical representations show the same numerator and the same denominator, they both show 1/2 (“ident-numerator” and “ident-denominator”). In establishing such connections between multiple graphical representations, the student forms an abstract mental model of what a fraction is: some quantity defined in relation to another quantity.

In addition to understanding fractions representations, being fluent with fractions, in representing fractions, and in relating different representations of fractions has been recognized as an important foundation of algebra learning (NMAP, 2008). Representational fluency means that students associate the abstract fraction shown by each graphical representation without reasoning about the knowledge components of numerator, denominator, and the unit separately. Instead, they treat each graphical representation as one perceptual chunk that stands for a given fraction. In other words, the student does not have to rely on the separate knowledge components “ident-numerator-circle” and “ident-denominator-circle”, but acquires a new knowledge component of “ident-fract-circle”. This knowledge component integrates the perceptual properties of the given graphical representation, thus enabling the student to estimate the fraction a circle shows (e.g., a circle shows about 1/2 or about 3/4). Being fluent in using individual graphical representations frees cognitive capacities students can invest in other learning processes and thereby enhances future learning.

Connectional fluency allows students to “simply see” that different graphical representations show the same fraction, without having to reason about their equivalence based on corresponding knowledge components. That is, instead of having separate knowledge components for finding the corresponding elements of numerator and denominator in circles and number lines (“connect-numerator-circle-nl” and “connect-denominator-circle-nl”), the student acquires a new knowledge component that represents a perceptual chunk (“connect-circle-nl”) and enables students to identify circles and number lines that show the same fraction by making use of their perceptual properties. The ability to fluently make connections between graphical representations allows students to work flexibly with them and frees cognitive capacities that students can invest in other learning activities that require flexible use of a variety of graphical representations.

The theoretical framework just described crucially informed the instructional design of the Fractions Tutor (see section 3), which serves as the research platform for my experimental studies (see section 4). Furthermore, I discuss the results from the experimental studies in the light of the theoretical framework.

3 Fractions Tutor

In this section, I describe a further contribution of my dissertation work: the Fractions Tutor. The Fractions Tutor is a successful intelligent tutoring system that promotes students' robust learning of fractions and is usable within the context of real classroom settings.

I first describe my motivation in developing a multi-representational tutoring system for the domain of fractions. I then describe in detail the way in which the Fractions Tutor incorporates interactive graphical representations, and how its use of graphical representations differs from other existing intelligent tutoring systems and other educational technologies. Next, I describe the use of user-centered methods to inform specific design decisions throughout the development process. Then, I describe the curriculum and design of the Fractions Tutor in detail. Finally, I briefly discuss empirical findings on the effectiveness of the Fractions Tutor.

3.1 Multiple graphical representations to help students learn fractions

Understanding fractions is foundational for learning algebra and more advanced math (NMAP, 2008; Siegler et al., 2012). Yet, fractions pose a significant challenge for students in the elementary and middle grades, and even for college students and pre-service teachers (Kaminski, 2002; Person, Berenson, & Greenspon, 2004). For example, the average 4th-grade student performed only at the basic level in the 2011 national NAEP math assessment, which included fractions and rational numbers (http://nces.ed.gov/nationsreportcard/). Indeed, fractions is the point when math often stops making sense to children (Moss, 2005). The difficulties that young students have with fractions are well documented (Boyer, Levine, & Huttenlocher, 2008; Brinker, 1997; Callingham & Watson, 2004; Person et al., 2004; Pitta-Pantazi, Gray, & Christou, 2004; Riddle & Rodzwell, 2000; Steencken & Maher, 2002; Tatsuoka, 1984; Tunç-Pekkan, Zeylikman, & Rummel, 2010; Witherspoon, 1993).

Students' problems with fractions may not come as a surprise: fractions are a complex topic (Charalambous & Pitta-Pantazi, 2007; Meagher, 2002; Ohlsson, 1991; Paik, 2005; Post, Behr, & Lesh, 1982) that involves counting (Fuson, 1988), proportional and multiplicative reasoning (Boyer et al., 2008; Hecht, Vagi, Torgesen, Berch, & Mazzocco, 2007; Kent, Arnosky, & McMonagle, 2002; Kieren, 1993; Lesh, Post, & Behr, 1988; Post et al., 1982; Stafylidou & Vosniadou, 2004; Thompson & Saldanha, 2003; Vanhille & Baroody, 2002) in a way that fundamentally differs from students' prior experience with whole numbers (Mack, 1995; Mack, 1993; Ni & Zhou, 2005). Fractions encompass complex concepts involving measurement (Carpenter, 1971), ratios (Fuson & Abrahamson, 2005; Lamon, 1999), equi-partitioning (Confrey & Maloney, 2010), units and re-unitizing (Cramer & Henry, 2002; Cramer & Wyberg, 2009; Lamon, 1999; Tunc-Pekkan, Rau, Aleven, & Rummel, 2010; Yanik, Helding, & Flores, 2008). Students need to understand relations between fractions and concepts familiar from out-of-school contexts (Ball, 1990), with whole numbers, decimals and other rational numbers (Behr, Post, Harel, & Lesh, 1993; Cramer, Wyberg, & Leavitt, 2009; Pagni, 2004; Siegler et al., 2011). Students need to learn how to perform operations such as adding and subtracting fractions (Cramer et al., 2008; Test & Ellis, 2005; Torbeyns, Verschaffel, & Ghesquière, 2005), finding equivalent fractions (Kamii & Clark, 1995; Ni, 2001), and multiplying fractions (Taber, 2001). Students need to perform mental computations (Caney & Watson, 2003; Steiner & Stoecklin, 1997), and estimate relative magnitudes (Caney & Watson, 2003; Cramer & Wyberg, 2007). Understanding fractions involves a complex set of mappings between real-world scenarios, representations, and symbols (Paik, 2005). In summary, there are many different conceptual interpretations of fractions (Charalambous & Pitta-Pantazi, 2007; Kieren, 1993): fractions can be understood as parts of a whole, as measurements, or as ratios. A large part of the difficulty that students have in understanding fractions is related to understanding these different conceptual interpretations and relating them to one another.

In fractions instruction, graphical representations are often used to help students understand these different conceptual interpretations of fractions (Charalambous & Pitta-Pantazi, 2007; Kieren, 1993; Lamon, 1999; Martinie & Bay-Williams, 2003; Moss & Case, 1999; Thompson & Saldanha, 2003). Commonly used graphical representations of fractions include area models (e.g., fraction circles, geoboards), linear models (e.g., fraction strips, cuisenaire rods, number lines), and discrete models (e.g., sets, counters). In fact, understanding and coordinating between different graphical representations between them is regarded key to students' success in understanding fractions (NCTM, 2000, 2006; NMAP, 2008; Siegler et al., 2010; Thompson & Saldanha, 2003). Several observational studies show that providing instruction that helps students relate these graphical representations to underlying concepts of fractions can promote learning (Brinker, 1997; Corwin et al., 1990; Cramer & Wyberg, 2009; Cramer et al., 2008; Mack, 1995; Moss, 2005; Paik, 2005; Pitta-Pantazi et al., 2004; Yang & Reys, 2001). Furthermore, helping students make connections between different representations of fractions has been shown to be effective in observational studies (Moss, 2005; Moss & Case, 1999; Taber, 2001) and in case studies (Kafai, Franke, Ching, & Shih, 1998). In my own prior work (see section 4.1), I provide experimental evidence that students working with multiple graphical representations of fractions learn better than students who work with only a single graphical representation, although only when prompted to explain how the graphical representations (e.g., half of a circle) of fractions relate to the symbolic representation (e.g., 1/2; Rau et al., 2009). This study demonstrates that we cannot take students' benefit from multiple graphical representations for granted; rather, whether or not multiple graphical representations are helpful depends on the kind of instructional support students receive to learn from them.

My goal in developing the Fractions Tutor was therefore to use multiple graphical representations with appropriate instructional support in a way that enhances students’ robust learning of fractions knowledge.

3.2 Use of multiple, interactive, abstract graphical representations

The Fractions Tutor includes several *abstract* and *interactive* graphical representations: circle diagrams, rectangles, and number lines (Fig. 5). Each graphical representation emphasizes certain aspects of different conceptual interpretations of fractions (Charalambous & Pitta-Pantazi, 2007). The circle as a part-whole representation depicts fractions as parts of an area that is partitioned into equally-sized sections. The rectangle is a more elaborate part-whole representation as it can be partitioned vertically and horizontally. Unlike the circle, the rectangle does not have a standard shape for the unit, thus allowing for more flexible re-unitizing procedures. Finally, the number line is considered a measurement representation and thus emphasizes that fractions can be compared in terms of their magnitude, and that they fall between whole numbers.

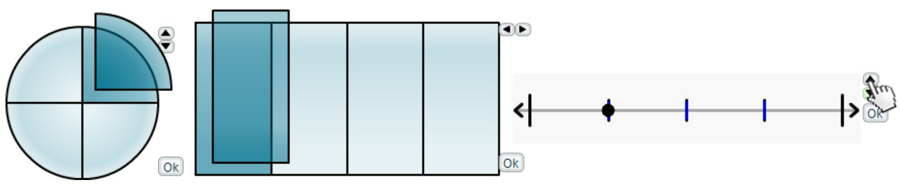


Fig. 5. Interactive circle, rectangle, and number line representations, as used in the Fractions Tutor.

Fig. 6 shows an example problem of the Fractions Tutor in which students use an interactive number line to find the location of “1” given a dot that shows 1/6. In the example problem in Fig. 7, students learn about addition using an interactive rectangle.

The Fractions Tutor includes *abstract* graphical representations based on the notion that they lead to more transferable knowledge because the representation is not tied to a specific scenario (Goldstone & Son, 2005; Goldstone, Steyvers, & Rogosky, 2003; Smith, 2003). In addition, abstract representations may be advantageous because they facilitate interpretations of a situation in terms of abstract relations rather than specific attributes (Resnick & Omanson, 1987; Schwartz & Black, 1996). However, to promote students’ understanding of graphical representations based on their prior real-world experiences (Grady, 1998; Heim, 2000; Nisbett & Ross, 1980), the Fractions Tutor introduces the abstract graphical representations within real-world contexts and concrete representations (e.g., pizzas, chocolate bars). This approach of using abstract graphical representations while introducing them with concrete graphical representations corresponds to Goldstone and Son’s (2005) approach of “concreteness fading”, which was shown to be successful in an experimental study. Comparisons across several iterations of classroom studies (descri-

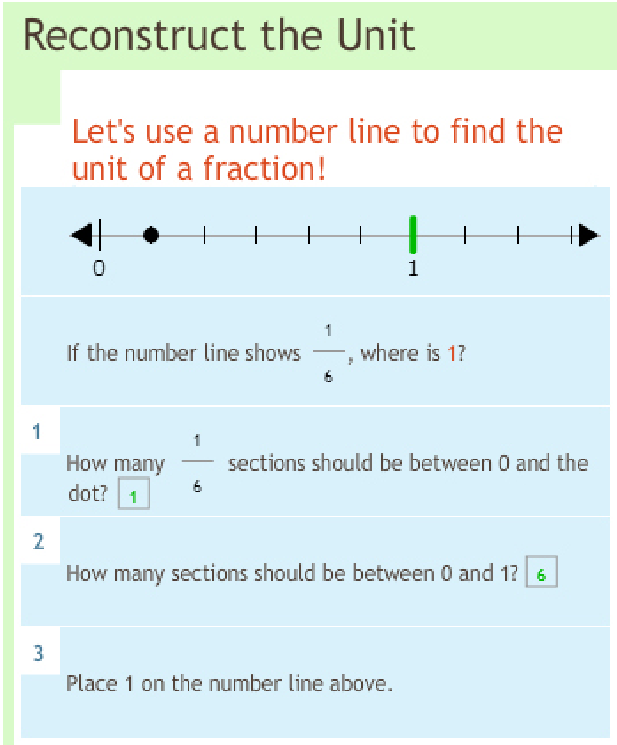


Fig. 6. Example problem of the Fractions Tutor in which students use an interactive number line to find “1” given a dot that shows 1/6.

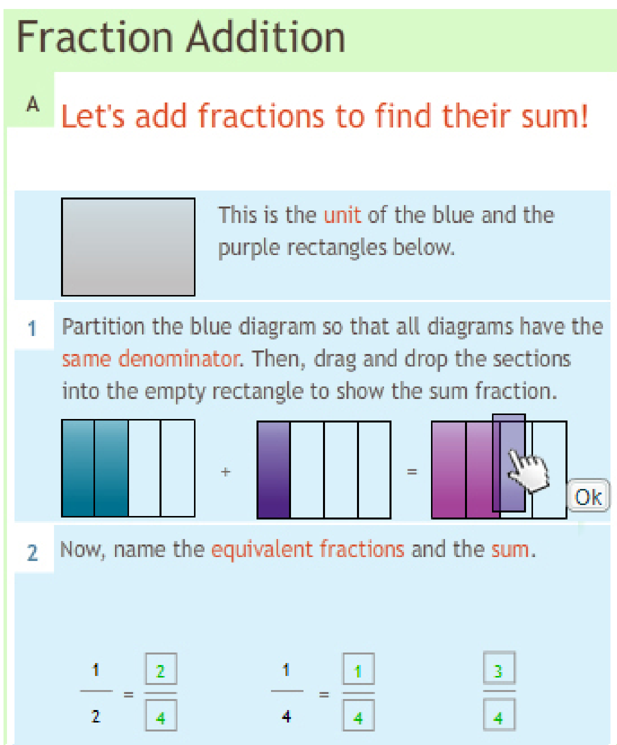


Fig. 7. Example problem of the Fractions Tutor in which students use an interactive rectangle to add two given fractions.

bed below) provide some (albeit non-experimental) evidence that students enjoy a version of the Fractions Tutor more if it includes problems that introduce the abstract graphical representations in the context of realistic scenarios rather than purely abstract representations without realistic contexts (Rau, Aleven, Rummel et al., 2013).

The Fractions Tutor uses *interactive* graphical representations that students can manipulate and use as tools to solve fractions problems. Several studies have discussed advantages of using virtual manipulatives, or interactive graphical representations, to help students understand fractions (Durmus & Karakirik, 2006; Kafai et al., 1998; Moyer et al., 2002; Proctor et al., 2002; Reimer & Moyer, 2005). The use of interactive graphical representations is based on the observation that physical activities, such as paper folding (Kamii & Clark, 1995), or the use of physical manipulatives (Caldwell, 1995; Cramer & Henry, 2002; Martin, Svihla, & Smith, 2012; Moss, 2005; Moss & Case, 1999) promotes students' learning of fractions. Reimer and Moyer conducted an observational study in a classroom of 3rd-grade students and find that virtual manipulatives have some advantages over physical manipulatives, such as allowing for more immediate and specific feedback, easier and faster interactions, and increased student enjoyment (Reimer & Moyer, 2005). Suh and colleagues demonstrate the effectiveness of virtual manipulatives in an observational study in 5th-grade classrooms (Suh et al., 2005). In a quasi-experiment on proportional reasoning with 3rd-grade students, they demonstrate that virtual manipulatives are as effective as physical manipulatives (Suh & Moyer, 2007). In a classroom experiment on fractions learning, Roussou and colleagues compared the effectiveness of supporting fractions learning with virtual building blocks to physical building blocks (Roussou et al., 2006). While they do not find differences between conditions, they discuss several advantages of learning with virtual building blocks based on qualitative analyses of individual cases. Proctor and colleagues (Proctor et al., 2002) describe a case study in which virtual manipulatives were used for remedial instruction to help one 7th-grade students' understanding of fractions. Lamberty and Kolodner (2002) describe a case study in which students use a virtual quilt tool to learn about fractions. Yet, in none of these studies were interactive graphical representations used within an intelligent tutoring system to address students' misconceptions of fractions by providing adaptive feedback on their interactions with graphical representations. In the light of this research, the Fractions Tutor uses interactive graphical representations while providing hints on demand and immediate feedback on each interaction with the representations. The Fractions Tutor detects commonly made errors on these interactions (e.g., dividing a circle diagram into 4 sections rather than 5 to show the fraction 4/5) and provides individualized feedback to remedy common misconceptions that become evident in students' interactions with graphical representations.

Many other intelligent tutoring systems include interactive graphical representations. The focus of the Fractions Tutor is novel in that it supports *conceptual* learning with *multiple, interactive, abstract* graphical representations. ASSISTments, a system for middle-school math (Heffernan, Heffernan, Deceoteu, & Militello, 2012), focuses on procedural rather than conceptual tasks. ActiveMath, an intelligent tutoring system that supports self-regulated learning of fractions based on a constructivist approach (Goguadze, Melis, & DFKI, 2008), includes mainly non-interactive graphical representations that update in response to changes students make in corresponding symbolic fractions. Kong and colleagues (Kong, 2008; Kong & Kwog, 2003; Kong, Lam, & Kwog, 2005) describe an intelligent tutoring system for fractions that relies on rectangle representations only. In Animalwatch (Beal, Arroyo, Cohen, Woolf, & Beal, 2010), students interact with various concrete graphical representations of fractions (e.g., sets of dogs, lengths of buttons), but it does not include abstract graphical representations. Many other learning environments for fractions exist that use multiple, interactive, abstract graphical representations (Adauto & Klein, 2010; Akpinar & Hartley, 1996; Kafai et al., 1998; Reimer & Moyer, 2005), but these are not intelligent tutoring systems and hence do not provide adaptive feedback and hints on demand on students' interactions with graphical representations.

In sum, the Fractions Tutor is unique in its use of multiple abstract, interactive, graphical representations on which students receive individualized feedback to address their misconceptions about fractions.

3.3 Principled approach to resolving stakeholder conflicts

In developing the Fractions Tutor, I took a user-centered design approach that takes into account stakeholder goals and needs. The objective was to develop a system that not only promotes robust learning of fractions knowledge, but that is also usable within the context of real classrooms. It is almost impossible to develop a system that satisfies all stakeholders, as students and teachers (among others) often have different goals. For example, students may want to work with a fun and enjoyable system, whereas teachers want students to learn and perform well on standardized tests, while also keeping the classroom under control. Though these goals and needs do not necessarily conflict, conflicts inevitably arise in the complex context of real educational settings, for instance between students' need for fun and entertainment and teachers' goal of classroom management. When they do arise, it is difficult to weigh them against one another. Usually, developers of educational technologies have to rely on ad-hoc or intuitive methods to address such conflicts. To address the conflicts that I faced when developing the Fractions Tutor, I developed a new methodology that integrates methods from human-computer interaction, intelligent tutoring systems, and learning sciences.

Since I identified conflicts between stakeholder goals throughout the design process of the Fractions Tutor, I first review how I applied the development process for Cognitive Tutors (Corbett et al., 2001; Koedinger, 2002; Koedinger & Corbett, 2006) to the design of the Fractions Tutor. Then I describe my approach to resolve conflicts between stakeholder goals in the Fractions Tutor.

3.3.1 Cognitive Tutor design process

The design process for Cognitive Tutors comprises a set of iterative, non-linear stages.

3.3.1.1. Stage 1: Stakeholder and problem identification

The first step in Cognitive Tutor design is to identify the educational problem to be addressed as well as stakeholders and their objectives. To accomplish this goal, I interviewed students, teachers and curriculum developers, review education literature, national and state standards.

As described above (see section 3.1), the development of the Fractions Tutor was motivated by the fact that students struggle with fractions as early as elementary school (NMAP, 2008; Siegler et al., 2010) although fractions are considered an important educational goal (Siegler et al., 2012). At the beginning of the development process, I conducted semi-structured interviews with five middle-school teachers, each with over 10 years of experience in teaching fractions. Teachers were asked to describe their instructional strategy for introducing fractions, their use of graphical representations, and what functionalities educational software for fractions learning should provide. These interviews demonstrated that teachers typically use a variety of graphical representations. They stated that educational software that uses interactive graphical representations while encouraging students to make connections between the different graphical representations and between the graphical and the symbolical representations might help students overcome their difficulties with fractions.

3.3.1.2. Stage 2: Identifying assessment and practice problems

Based on the educational problem, I identified a set of assessment tasks (i.e., tasks learners should be able to solve after having worked with the educational software). These assessment tasks guided the selection of practice problems (i.e., problems students should solve as part of the educational software). A search of the education literature yielded a set of domain-specific target problems both for assessment and for practice. In addition, I brainstormed with research group members and teachers about novel problems for assessment and practice.

The outcome of stage 2 was a set of domain-specific assessment tasks and practice problems for the Fractions Tutor. Along with the Fractions Tutor, the assessment tasks went through a number of iterations that reflected alignment with the practice problems, and were repeatedly validated based on reliability and validity analyses.

3.3.1.3. Stage 3: Cognitive task analysis

The goal of stage 3 is to understand student learning and student thinking in the domain. In doing so, I identified the knowledge and strategies the Fractions Tutor should cover using cognitive task analysis techniques (Clark, Feldon, van Merriënboer, Yates, & Early, 2007; Koedinger et al., 2012; Schraagen, Chipman, & Shalin, 2000). Cognitive task analysis seeks to identify the knowledge components that students need to acquire to perform well on assessment and practice problems. Cognitive task analysis can employ think-aloud protocols and observations of student learners (novices) or proficient student (experts), or make use of a theory of what knowledge learners need to acquire. In developing the Fractions Tutor, I combined think-aloud protocols and observations with difficulty factors assessment (Baker, Corbett, & Koedinger, 2007; Koedinger et al., 2012; Koedinger & Nathan, 2004) – a method to identify features of tasks that reliably change the difficulty of the task.

I conducted several iterations between stages 2 and 3. After each iteration, I updated the collection of assessment and practice problems based on insights gained from cognitive task analysis and difficulty factors assessments. I reviewed the problems in focus groups with research group members and teachers. As part of this stage, I also discussed potential problem sequences for the Fractions Tutor with other researchers and teachers. In addition, I used observations of teacher-student tutoring to guide the instructional design, which provided insights into successful instructional strategies.

The outcome of stage 3 was a set of knowledge components and of practice problems that address all knowledge components in order of ascending difficulty.

3.3.1.4. Stage 4: Cognitive modeling and tutor development

Stage 4 aims at developing the Cognitive Tutor. As part of this stage, I created the Fractions Tutor interface, a cognitive model of student problem solving that serves as a basis for individualized support, and a curriculum that contains a collection of problem types and that span across a variety of topics that the Fractions Tutor covers.

Stage 4 included several cycles of rapid, low-fidelity prototyping and high-fidelity prototyping. These rounds of testing were conducted in the laboratory with a small number of students from the target population. Between each round of testing, I updated the materials based on the findings and issues identified. The Fractions Tutor is an example-tracing tutor that I developed using Cognitive Tutor Authoring Tools (CTAT, Aleven, 2010; Aleven, McLaren, Sewall, & Koedinger, 2006; Aleven et al., 2009a). CTAT allows for rapid prototyping and fast implementation of iterative design changes. I included teachers in the development process by asking them to repeatedly review and comment on initial designs and early prototypes. Involving teachers not only improved the quality of the educational software – it also helped establish relations with teachers and revealed further stakeholder goals.

The outcome of stage 4 is a set of multiple fully-functioning versions of the Fractions Tutor ready for further testing and evaluation.

3.3.1.5. Stage 5: Pilot studies and parametric studies

The goal of stage 5 was to formally evaluate and iteratively improve the Fractions Tutor. I made use of a variety of methods during this phase. First, pilot testing the Fractions Tutor in the laboratory was useful to get in-depth insights with students solving practice problems while thinking aloud, which helped identify gaps in their knowledge that the Fractions Tutor did not yet address. Second, testing the educational software in classrooms is indispensible. I gathered a variety of data from classroom studies with the Fractions Tutor. I assessed students' learning gains based on pretests and posttests that integrate the target problems identified during earlier stages, including both standardized test items and transfer items that assess students' ability to apply their knowledge to novel task types. In addition, informal observational data of students' interactions with the Fractions Tutor in classrooms, interviews and focus groups with teachers as well as surveys with students and with teachers yielded valuable insights into usability issues and reveal crucial aspects of the stakeholders' goals. Finally, log data gathered while students used the Fractions Tutor provided a useful basis for identifying issues in usability and difficulty level of particular steps within the educational software.

As part of this stage, I conducted a series of parametric studies in classrooms (described in detail in section 4), which investigated how best to implement multiple graphical representations so they enhance robust learning of domain knowledge. These studies led to iterative improvement of the Fractions Tutor. As a consequence, the Fractions Tutor uses graphical representations in the following manner:

- The Fractions Tutor encourages students to reflect on the relation between each graphical representations (e.g., circles) and the corresponding symbolic representation (e.g., ½) in the form of menu-based prompts (see Experiment 1, section 4.1, Rau et al., 2009).
- The Fractions Tutor uses a spiral curriculum that switches frequently between different topics (e.g., equivalent fractions, fraction addition; see Experiment 2, section 4.2, Rau, Aleven et al., 2013b).
- The Fractions Tutor frequently switches between different graphical representations (see Experiment 3, section 4.3, Rau, Rummel et al., 2012).
- The Fractions Tutor provides connectional sense-making support by encouraging students to become active in making sense of connections between different graphical representations

(e.g., circles and number lines) through the use of worked examples (see Experiment 4, section 4.4, Rau, Aleven, & Rummel, 2013a; Rau, Aleven et al., 2012).

- The Fractions Tutor provides connectional fluency-building support by the means of mixed representation problems (see Experiment 4, section 4.4, Rau, Aleven, & Rummel, 2013a; Rau, Aleven et al., 2012)
- The Fractions Tutor provides connectional sense-making support before connectional fluency-building support (see Experiment 5, section 4.5, Rau, Aleven, & Rummel, 2013a).

The outcome of stage 5 was a set of updated stakeholder goals, as well as an updated and iteratively improved version of Fractions Tutor ready for classroom dissemination.

3.3.1.6. Stage 6: Classroom use and evaluation

The goal of stage 6 is to evaluate the educational software in the field. I conducted a number of randomized field trials during this phase. After several iterations, I evaluated the Fractions Tutor in classroom studies. I evaluated the Fractions Tutor not only based on students' performance on pretests and posttests. Observations in randomly selected classrooms, interviews with randomly selected teachers and student or teacher surveys further helped identifying problem-solving behaviors and learning processes. In addition, the analysis of tutor log data during problem solving served as a basis for detecting issues with specific problem-solving steps, for example by identifying steps on which students make many errors. I briefly describe the results from my most recent evaluation below (see section 3.5).

3.3.2 Identifying stakeholder goals and instructional design principles

I now describe a novel contribution to the design processes just described, which crucially informed the design of the Fractions Tutor. In doing so, I will focus on a few examples, which illustrate the use of my methodology. For a more complete description of the design process, please refer to Rau, Aleven, Rummel and colleagues (2013). I first describe a hierarchy of stakeholder goals that constitutes the basis of the design process. Then, I describe the instructional design recommendations that I identified to address these goals and identify conflicts between instructional design recommendations. Table 1 gives an overview of the instructional goals and instructional design recommendations as well as of the resulting design conflicts. Finally, I present three approaches to resolving these conflicts.

Priority	Goal	Instructional design principles	Design conflicts
1	Robust learning (G1)	Realistic cover stories (id1) Subgoaling (id2) Sparse use of color (id3) Holistic, complex problems (id4)	C1: Subgoaling (id2) vs. holistic problems (id4)
2	Classroom management (G2)	Easy problems (id5)	C2: realistic cover stories (id1) vs. easy problems (id5)
3	Entertainment (G3)	Colorful, flashy elements (id6)	C3: Sparse use of color (id3) vs. colorful, flashy elements (id6)

Table 1. Overview of example goals, instructional design principles, and design conflicts.

3.3.2.1. Forming a goal hierarchy

Across the stages and iterations of tutor development, I kept track of stakeholder goals. Based on focus groups and interviews with teachers and students, which I conducted as part of each stage, I created a goal hierarchy to identify and resolve goal conflicts.

To develop a goal hierarchy, I used affinity diagrams, a common human-computer interaction technique (Beyer & Holtzblatt, 1998): I wrote each goal on a sticky note and then worked bottom-up to organize them into a hierarchy. Once all notes were collected in groups, I named the group. I then identified a set of instructional design recommendations that could help achieve each goal.

Goals and instructional design recommendations

Based on interviews with teachers and based on the review of educational standards (Siegler et al., 2010; NMAP, 2008; NCTM, 2000), I identified teachers' goals to promote students' learning of robust knowledge about fractions, which can transfer to new problem types and that lasts over time (G1). Furthermore, I used education standards and math literature (as described in detail in section 3.1) to formulate domain-specific goals related to promoting conceptual understanding of fractions as parts of a whole, as proportions, and as measurements. The Fractions Tutor design incorporates many instructional design recommendations from the math education literature and the learning sciences literatures. One of these recommendations recommends using complex realistic problems with cover stories (id1; Bassok, 1996; Blessing & Ross, 1996; Nunes, Schliemann, & Carraher, 1993; Thußbas, 2001). Furthermore, the literature recommends to illustrate the structural components of a problem-solving procedure using subgoaling (id2). Subgoaling is a procedure that aims at communicating the goal structure of a problem by breaking it into clear substeps, thereby “making thinking visible” (Anderson, Corbett, Koedinger, & Pelletier, 1995; Catrambone & Holyoak, 1998; Singley, 1990). The Cognitive Theory of Multimedia Learning (Mayer, 2003; Mayer, 2005) suggests to use color only sparingly, so as to highlight only conceptually relevant aspects of the problem but without adding extraneous distracters (id3). Finally, constructivist learning theories (Bereiter & Scardamalia, 2003; Cobb, 1995; Kafai et al., 1998; Mintzes, Wandersee, & Novak, 1997) recommend that students work on complex, holistic problems (id4).

Classroom observations demonstrated teachers’ needs for classroom management while using the Fractions Tutor (G2), including the ability to focus on students who struggle with the content, monitoring students’ progress, and a quiet classroom of students who concentrate on their work. For instance, when asked what they like about using educational software such as the Fractions Tutor, teachers reported: “I like using it because it is so interactive for the students. They stay very involved,” or “The programs that I use with my students are interactive, colorful, and can hold their attention.” To address teachers’ goal for classroom management, I used focus groups with teachers to identify possible obstacles that the Fractions Tutor created within the classroom. I discovered that any aspect that makes the educational software difficult to use for students results in teachers helping students out with usability issues rather than helping with the content. A piece of educational software that is easy to use and that includes easy math problems would thus help achieve this goal (id5).

Finally, surveys with students demonstrated their goal to have fun and to be entertained (G3). This need might best be achieved focusing on age-appropriate design elements resembling games with colorful and flashy elements (id6). Also, complex real-world problems with cover stories might address students’ need for interesting practice problems (see id1).

Hierarchy of goal categories

Next, I created a hierarchy of the goal categories just described. In doing so, focus groups with the stakeholders informed the ranking of goals. For the goals on which I could not find consensus in focus groups, I again used affinity diagrams to identify classes of goals based on the effect they have on students’ learning and on the dissemination of the Fractions Tutor. In doing so, I conducted a brainstorming session with experts (who have good knowledge of the relevant literatures) about the effects of common interventions to meet the goals can help create the goal hierarchy. I then regrouped the generated items to create a diagram for the effects. Next, I computed an impact factor for each goal and ordered the goals accordingly. Goals that also served the attainment of other goals (e.g., increasing students' concentration through improved classroom management also promotes the goal to help students learn) were given a higher impact factor than goals that impeded another goal (e.g., including colorful but distracting elements may impede learning). Altogether, goals with a higher impact factor were given priority in resolving design conflicts. Table 1 gives an overview of the resulting hierarchy including the instructional design recommendations and resulting design conflicts.

Conflicts

I now turn to mapping out a few example conflicts that arise from competing goals and from the resulting instructional design recommendations just described. To identify these conflicts, I conducted focus groups with learning sciences experts who have in-depth knowledge of the empirical research on the various design recommendations.

Conflicts can arise between design recommendations that address the same goal. One such *conflict C1* exists within the goal to promote robust learning (G1) by using subgoaling to break the problem up into small steps (id2) or by providing holistic and complex problems (id4).

Further conflicts can arise from constraints within schools and students' abilities. For example, given students' reading ability is often poor, *conflict C2* occurs between the goal to promote robust learning (G1) by providing complex real-world problems with cover stories (id1) and teachers' needs to facilitate classroom management (G2) by providing an easy-to-use system (id5): in my own classroom studies, I found that the increased reading effort due to the use of cover stories was impractical in classrooms given students' low reading ability. Teachers were busy helping students understand problem statements, rather than helping them with the math problems.

Finally, conflicts can arise between design recommendations on how to promote robust learning (G1) and students' emotional needs for fun and entertainment (G3). One such *conflict C3* exists between the use of lean designs so as to not distract the user from the learning task by using colorful highlighting only sparingly to emphasize conceptually relevant aspects (id3) and students' preference for flashy designs recommend the inclusion of game-like elements whose main purpose is to visually appeal to young students (id6).

3.3.3 Resolving conflicts

To address these conflicts in a principled way, I used three approaches: (1) where possible, I resolved conflicts based on the goal hierarchy, (2) I conducted parametric experiments, and (3) I conducted cross-iteration studies. Although I present these three approaches as a sequence, they complement each other and can occur at any of the iterative stages in the Cognitive Tutor design process described above (see section 3.3.1).

3.3.3.1. Goal hierarchy

To illustrate how I resolved conflicts based on the goal hierarchy, let us again consider conflict C3 between robust learning (G1) and students' goal to have fun (G3; i.e., inclusion of game-like, colorful elements whose main purpose is to visually appeal young students [id6] versus lean designs that use colorful highlighting to emphasize conceptually relevant aspects [id3]). Based on expert interviews and focus groups, the goal hierarchy places the highest priority on supporting robust learning (G1), whereas the goal to have fun (G3) has lowest priority. Therefore, it is clear that the Fractions Tutor should prioritize on employing color-based highlighting only conceptually relevant aspects. However, this can be done in a way that is visually appealing to students of the target age group. Further, I integrated flashy and exciting elements where (or when) they do not distract, for instance, at the end of a practice problem.

Fig. 8 illustrates several key aspects of the solution I chose for the Fractions Tutor. First, the choice in color reflects the finding that students in grades 4 and 5 have a preference for less intense colors with lower saturation and value, compared to younger students (Jakobsdottir, Krey, & Sales, 1994; Pett & Wilson, 1996). I also made sure the colors I selected are gender neutral (Jakobsdottir et al., 1994). Second, in the service of using color to emphasize only conceptually relevant aspects, I use orange to highlight key words in each problem step. Finally, Fig. 8 shows a success message that the Fractions Tutor displays at the end of a problem. The message contains a short movie clip that flies in.

Across different problems, the Fractions Tutor provides a variety of different success messages. Data from a survey that 429 students filled out after working with the Fractions Tutor shows that students found it visually appealing. To the question whether they liked the layout and color choice of the interface, 61% of students responded “Yes, a lot!”, 28% responded “I don’t care,” and only 12% responded “No, not at all!” The difference between these response options was statistically significant, χ^2 (2, N = 429) = 236.86, $p < .001$.

For additional examples of goal conflicts that I resolved based on the goal hierarchy, please refer to Rau, Aleven, Rummel et al., 2013.

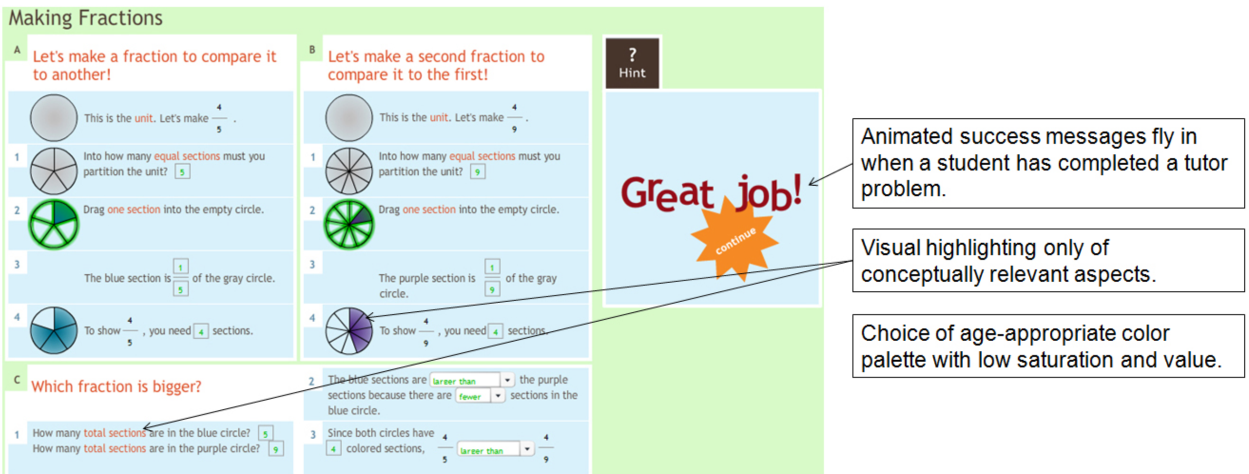


Fig. 8. Example problem in the Fractions Tutor that illustrates the use of animated success messages, sparse visual highlighting, and age-appropriate color palette.

3.3.3.2. Parametric experiments

Conflicts that cannot be resolved based on the goal hierarchy require more careful inspection. In this case, I conducted parametric experiments using multiple metrics to address important remaining conflicts. Consider again conflict C1 between subgoaling (id2) and holistic problems (id4). As mentioned, the subgoaling strategy breaks up problems into their substeps, in order to communicate the problem’s goal structure (Anderson et al., 1995; Catrambone & Holyoak, 1998; Singley, 1990). However, surveys with students who participated in a classroom experiment with 311 students indicated that students tend to dislike multi-step problems. A student commented, for example: “suggestions i would make is stop the repeating and give more fun stuff because i heard from people even me not to be mean but most of it ws boring sorry.” Another student said: “in my opinion that there were too many questions in one problem!!” Having many steps within a problem seems to overwhelm students. For example, a student reported: “I think there was too many questions.” To address this issue, I conducted a classroom experiment that investigated the incorporation of more holistic problems (see fluency-building problems described in section 3.4.3), in addition to problems that employ the subgoaling strategy (see single-representation problems and worked-example problems described in sections 3.4.1 and 3.4.2). Although this experiment included a number of variations and thus does not a tightly controlled test of a holistic approach versus a subgoaling approach, the results suggest that a version of the Fractions Tutor that includes 25% of holistic problems lead to better learning than a version with 50% of holistic problems while being more enjoyable than a version without any holistic problems. I describe this experiment in detail below (see Experiment 4, section 4.4; also see Rau et al., 2013).

3.3.3.3. Cross-iteration studies

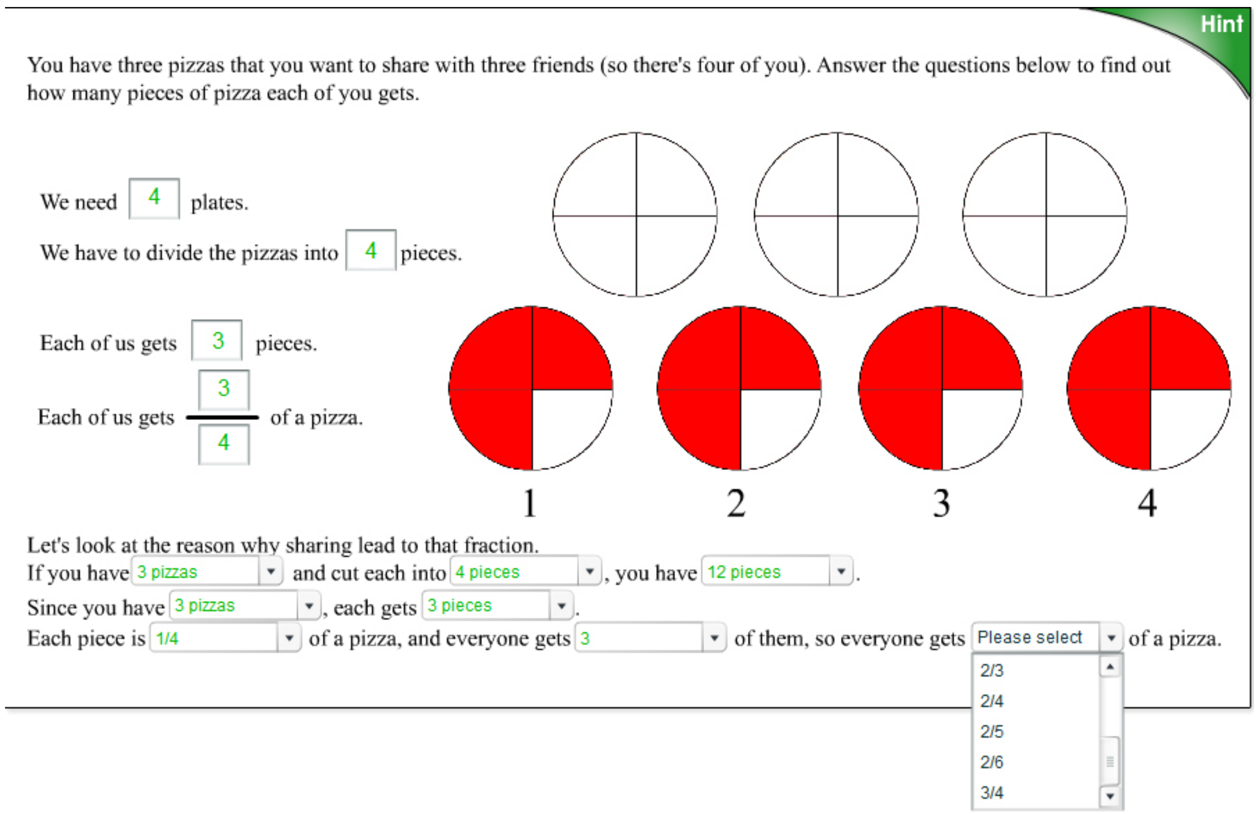


Fig. 9. Example of an early version of the Fractions Tutor with cover stories. Students learn about fractions in the context of sharing pizzas.

Unfortunately, it is not possible to conduct a controlled experiment for every design decision. In this case, I recommend conducting cross-iteration studies – to the extent possible (since it is not possible to iterate on every design decision either). For example, I addressed conflict C2 between the goal to promote robust learning (G1) by providing complex real-world problems with cover stories (id1) and the goal to facilitate classroom management (G2) by providing an easy-to-use system (id5) based on the effects of the design decision across several iterations of the Fractions Tutor.

Initially, I resolved conflict C2 based on the goal hierarchy, which prioritizes robust learning. Fig. 9 shows an example of an early version of the Fractions Tutor, which includes cover stories. However, when employing a version of the Fractions Tutor that included cover stories in classrooms, I faced challenging issues. Students complained about having to read a lot, and teachers expressed their concern about being able to use the Fractions Tutor in their classrooms without extra help. Several teachers suggested including an audio function, so that students could listen to the problem statement via headphones. However, since many schools lack the necessary equipment (i.e., headphones), I discarded that idea. Instead, I excluded cover stories from the Fractions Tutor. Fig. 10 shows an example of the next iteration of the Fractions Tutor, without cover stories.

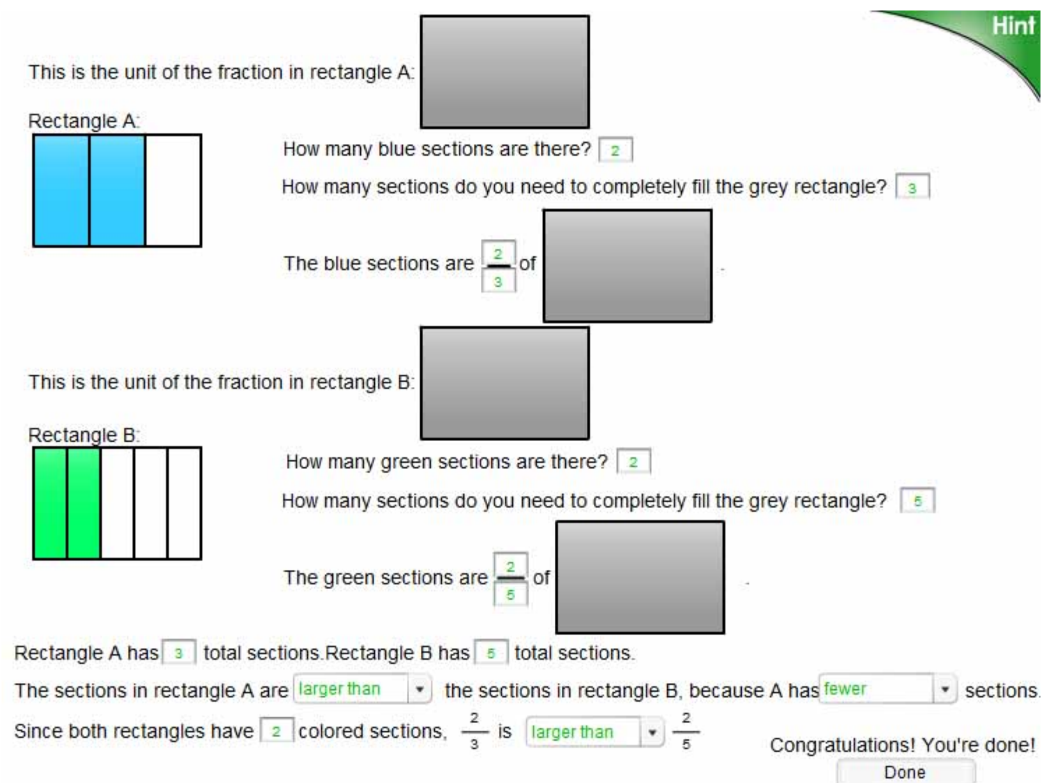


Fig. 10. Example of the next iteration of the Fractions Tutor without cover stories.

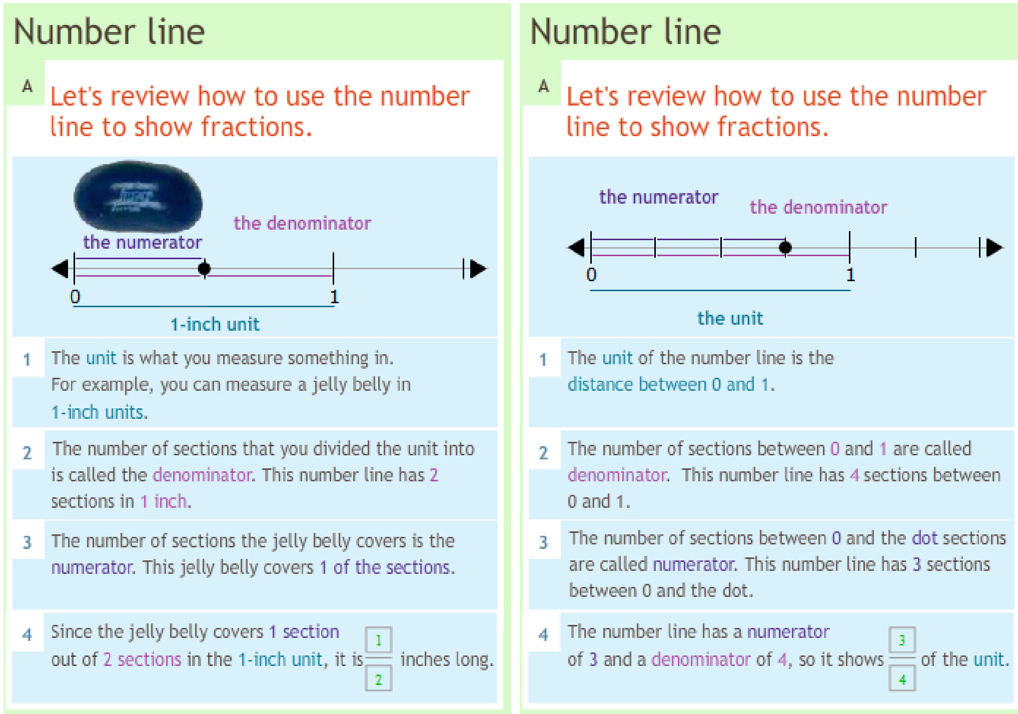


Fig. 11. Introductory problems with and without cover stories to introduce how the number line depicts fractions.

However, in a subsequent experiment, classroom observations demonstrated that students had trouble making sense of the rather abstract problems in the tutor. An anonymous survey with 331 students revealed that students thought the problems were too hard and that they were not fun. One student commented, for instance: "I don't like how the problem didn't give clear, vivid questions. It confused the way I was taught." Several students commented on the Fractions Tutor being boring, for instance: "it was good but it got boring at times."

I therefore included introductory problems that introduced the graphical representations used in the Fractions Tutor based on realistic cover stories (e.g., introducing number lines in the context of measuring the length of candy, see Fig. 11). The next round of classroom testing with a new version of the Fractions Tutor did not reveal any persisting issues with reading levels or the abstract language the Fractions Tutor uses. An anonymous survey with 429 students revealed generally positive comments. One student responded, for example: "fractions tutor is a really good learning program.the reason i like it was because it wasnt too hard and wasnt too easy. it was just right for me.also i learn alot just from this." Many students reported that they had fun with the tutor, for example: "i like about it is fun it makes people smart it was a lot fun."

These cross-iteration changes to the Fractions Tutor illustrate that in cases where design choice based on the goal hierarchy proves to be impractical, several iterations may be necessary to find a balance between unintended disadvantages of a desired design choice and alternative solutions. By carefully monitoring the effect of each design choice, I arrived at combining cover stories in introductory problems with less reading-intensive, abstract problems throughout the rest of the Fractions Tutor. Empirical findings from a sequence of subsequent classroom studies demonstrate that this choice is an effective and practical solution for the young population the Fractions Tutor is designed for.

One caveat of this approach is, however, that it is practically not possible to iterate on all design decisions, given that the design space of any educational software is very large. A further limitation of the present cross-iteration studies is that the changes that were made on the cover stories were confounded with other changes made in each iteration of the Fractions Tutor. It is difficult to rule out the possibility that these changes made no difference on the metrics I used to evaluate the tutor.

3.4 Overview of curriculum

Curricular Unit	Description
Topic 1	Naming unit fractions and proper fractions given graphical representations
Topic 2	Making graphical representations of unit fractions and proper fractions given symbolic
Topic 3	Reconstructing the unit of unit fractions and proper fractions
Topic 4	Naming improper fractions given graphical representations
Topic 5	Making graphical representations of improper fractions
Topic 6	Equivalent fractions: underlying concepts
Topic 7	Equivalent fractions: expanding and reducting
Topic 8	Comparing fractions
Topic 9	Adding fractions
Topic 10	Subtracting fractions

Table 2. Topics and problem types covered by the Fractions Tutor.

The Fractions Tutor curriculum covers ten topics (see Table 2), corresponding to over 10 hours of supplemental material. A more detailed description of the activities and concepts covered in each topic can be found in Appendix 1. The activities and concepts covered in each topic align with U.S. education standards including the Pennsylvania State standards, the NCTM standards, and the common core standards (see Appendices 2-4). A common theme throughout the Fractions Tutor is the unit of the fraction (i.e., what the fraction is taken of). The concept of the unit is introduced in the first topics. The concept of the unit lays the foundation for improper fractions (by demonstrating that fractions such as 1 1/2 can be larger than one unit), and to adding fractions (by showing that when adding fractions, the fraction is still defined with respect to the same unit). The focus on the unit throughout the curriculum illustrates the conceptual approach to problem solving that is characteristic of the Fractions Tutor.

The Fractions Tutor is designed for the use within classrooms. Students work on the tutor problems individually at their own pace, with a teacher present to help out individual students who need help. The Fractions Tutor is available for free at www.fractions.cs.cmu.edu through Mathtutor, a website for middle school math (Aleven, McLaren, & Sewall, 2009b). Students log onto the website with personal logins. Teachers are provided with a tool that allows them to retrieve information about their students' performance, for instance, how many errors students made on a particular set of problems. They can use this information to identify what problem type a particular student struggles with, and provide targeted advice to that individual student.

Based on the findings described in Experiment 2 (Rau et al., 2010; Rau, Aleven et al., 2013b, see section 4.2), the Fractions Tutor curriculum uses a spiral curriculum (Harden & Stamper, 1999): students work through the sequence of topics listed in Table 2 three times. For each topic, the Fractions Tutor includes problems with only one type of graphical representation (i.e., either a circle, a rectangle, or a number line), and problems in which multiple graphical representations are provided at the same time, paired with instructional support to make connections between them. In accordance with the findings described in Experiment 3 (Rau, Rummel et al., 2012, see section 4.3), the individual-representations problems are presented in interleaved fashion (i.e., one problem with a circle, followed by one problem with a rectangle, followed by one problem with a number line, etc.). Based on the findings described in Experiments 4 and 5 (Rau, Aleven et al., 2013a; Rau, Aleven et al., 2012, see sections 4.4 and 4.5), two different types of multiple-representations connection-making problems are presented after the individual-representation problems for each topic. In this section, I describe each of these different types of problems in detail.

3.4.1 Individual-representation problems for representational understanding and fluency

At the beginning of each topic in the Fractions Tutor, students are presented with problems that include only one graphical representation (i.e., either a circle, a rectangle, or a number line,). Fig. 12 shows a tutor problem that illustrates several of the features of the Fractions Tutor. Students are guided step by step through a fractions problem. Students interact with the circle representation by partitioning it into sections (see Fig. 12), by highlighting, and by dragging-and-dropping sections (see Fig. 5). The Fractions Tutor employs subgoaling (Anderson et al., 1995; Catrambone & Holyoak, 1998; Singley, 1990) to emphasize each knowledge component through visually separated steps. To focus students' attention to the step at hand, the tutor interface builds up step by step (e.g., step B-2 in Fig. 12 only shows up after the student has completed step B-1). Based on the findings described in Experiment 1 (Rau et al., 2009, see section 4.1), the Fractions Tutor provides menu-based reflection prompts (Aleven & Koedinger, 2000) at the end of each problem, for example to compare fractions by reasoning about the inverse relationship between the size of the denominator and the magnitude of the fraction (see bottom of Fig. 12).

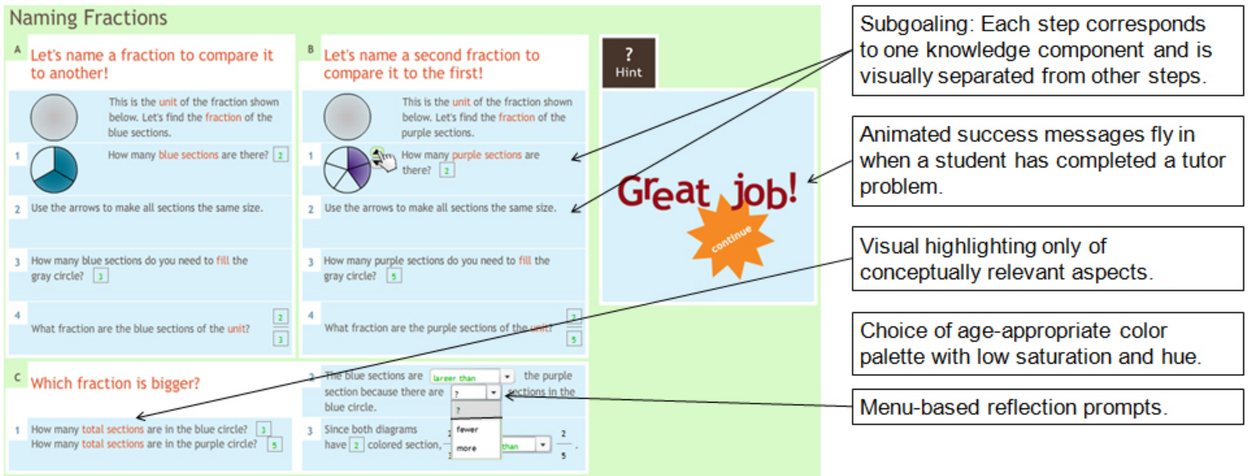


Fig. 12. Naming the fraction shown in an interactive circle, with reflection prompts.

3.4.2 Connectional sense-making support for connectional understanding

In addition to the individual-representation problems just described, the Fractions Tutor includes two main types of multi-representational tutor problems: problems with connectional sense-making support, and problems for connectional fluency-building support. Fig. 13 shows the screen shot of a connectional sense-making problem for equivalent fractions. Students learn that equivalent fractions in different graphical representations 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. 13). Worked examples have been shown to be an effective and efficient means to support students' learning in a variety of domains (Atkinson, Derry, Renkl, & Wortham, 2000; Atkinson & Renkl, 2007; Gerjets, Schwonke, & Catambone, 2006; Große & Renkl, 2007; Hilbert, Renkl, Kessler, & Reiss, 2008; Kopp, Stark, & Fischer, 2008; Kyun & Lee, 2009; Lewis & Barron, 2009; McLaren, Lim, & Koedinger, 2008; Nievelstein, van Gog, van Dijck, & Boshuizen, 2013; Nokes & VanLehn, 2008; Paas & van Gog, 2006; Paas & van Merriënboer, 1994; Pirolli & Anderson, 1985; Quilici & Mayer, 1996; Renkl, 1997; Renkl, 2002, 2005; Salden, Aleven, Renkl, & Schwonke, 2008; Schmidt-Weigand, Hänze, & Wodzinski, 2009; Schwonke et al., 2009; Schworm & Renkl, 2006; Sweller, 2006), including learning with multiple representations (Berthold et al., 2008; Berthold & Renkl, 2009; Schwonke, Berthold et al., 2009). Worked examples provide students with filled-in solution steps, which promotes learning by reducing cognitive load (Kirschner, 2002; Paas & van Gog, 2006; Paas & van Merriënboer, 1994; Renkl, Atkinson, & Große, 2004) because students do not have to invest mental effort into finding a solution through cognitively intensive strategies such as means-ends-search (Renkl, 2005). Instead, students can invest mental effort into making sense of the solution step (Renkl, 1997; Renkl, 2002). Several studies have successfully implemented worked examples in intelligent tutoring systems (Koedinger & Aleven, 2007; Salden et al., 2008; Schwonke, Renkl et al., 2009).

First, students are given worked example with a more familiar graphical representation: an area model representation (i.e., circle or rectangle). To ensure that students read through the worked examples, they are asked to fill in the last step themselves (step A-3 in Fig. 13). Next, once they complete that step, the problem-solving part of the worked example appears on the right (part B in Fig. 13). The problem-solving part always involves a more challenging graphical representation: the number line. Students can use the more familiar graphical representations from the worked example to guide their interactions with the more challenging graphical representation to solve a problem of the same type (e.g., finding equivalent fractions). 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. Thus, worked example problems help students make sense of the connections between representations by highlighting structural correspondences between a familiar and a less familiar graphical representation.

In the light of research showing that the positive effect of worked examples can be further enhanced by providing self-explanation prompts (Berthold & Renkl, 2009; Gerjets et al., 2006; Große & Renkl, 2007), students receive reflection prompts that help them abstract a general principle from the two graphical representations at the end of each worked example problem (part C in Fig. 13). For each step, students receive feedback at the level of the relevant knowledge components of the step at hand, for instance, by explaining that the denominator of a fraction corresponds to the number of total sections a rectangle is cut into.

Fig. 14 shows a screen shot of a connectional sense-making problem for fraction comparison. Connectional sense-making processes are 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 the support described for the equivalent fractions example (Fig. 13).

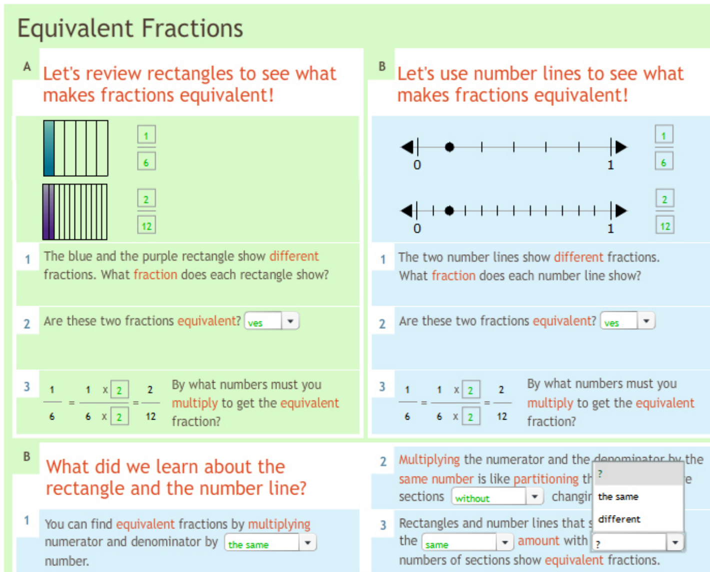


Fig. 13. Connectional sense-making support problem for equivalent fractions.

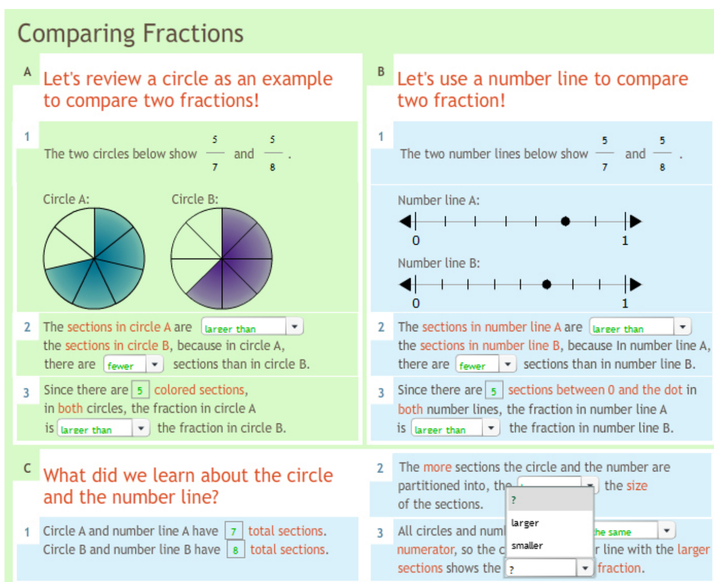


Fig. 14. Connectional sense-making support problem for fraction comparison.

3.4.3 Connectional fluency-building support for connectional fluency

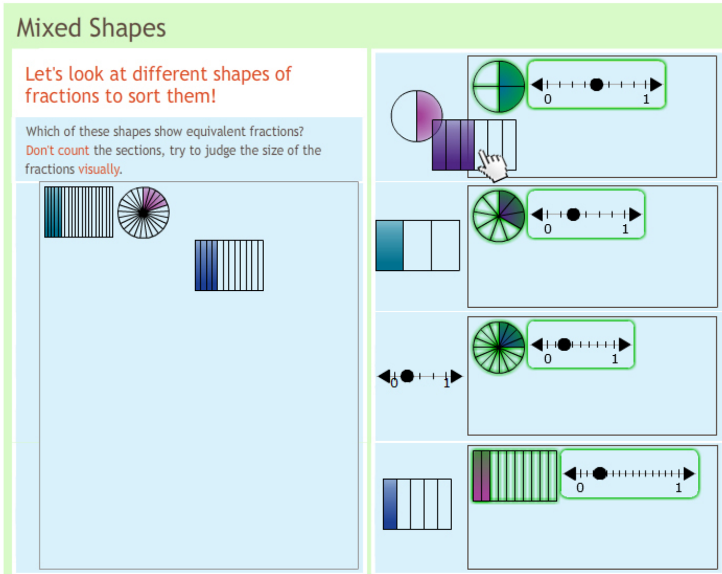


Fig. 15. Connectional fluency-building support problem for equivalent fractions.

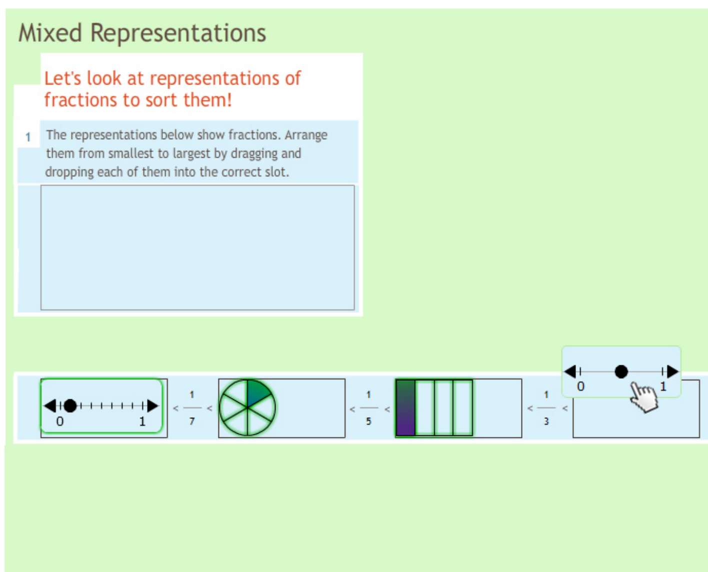


Fig. 16. Connectional fluency-building support problem for fraction comparison.

Fig. 15 shows a connectional fluency-building problem for equivalent fractions. The fluency-building problems are based on Kellman and colleague's perceptual learning paradigm (Kellman & Garrigan, 2009; Kellman et al., 2008; Kellman et al., 2009; Massey, Kellman, Roth, & Burke, in press). In these problems, students learn to relate different representations of math problems, such as graphical representations, text-based word problems, and symbolic representations, to one another based on their perceptual properties. Rather than making sense of *why* or *how* these different representations correspond to one another, connectional fluency-building problems aim at helping students in becoming *faster* and *more efficient* at extracting relevant information from the different representations based on repeated experience with a large variety of problems. Thus, connectional fluency-building problems help students make connections between representations fast and effortlessly through extensive perceptual experience.

In the Fractions Tutor's connectional fluency-building problems, students sort a variety of equivalent graphical representations using drag-and-drop. Rather than identifying numerator and denominator to solve the equivalence problem computationally, students visually judge whether graphical representations show equivalent fractions by estimating their relative size. As in Kellman and colleagues' fluency trainings (Kellman et al., 2008; Massey et al., in press), feedback is given only about the correctness of the sorting task, without referring to the underlying conceptual aspects, such as the knowledge components of numerator and denominator.

Fig. 16 shows an example of a connectional fluency-building problem for fraction comparison. Students sort graphical representations based on their relative size, using drag-and-drop. Again, students are encouraged to visually estimate the relative size of a variety of graphical representations.

3.5 Effectiveness

I evaluated the effectiveness of the Fractions Tutor in a sequence of classroom experiments (described in detail in section 4). Each experiment also served to iteratively improve the Fractions Tutor while investigating research questions following from the theoretical framework described above.

Results from the most recent classroom experiment (Rau, Aleven et al., 2012; see Experiment 4, section 4.4) showed that the Fractions Tutor leads to substantial learning gains that last over time. In this classroom experiment, 599 4th- and 5th-graders worked with the Fractions Tutor for 10 hours of their regular math instruction. Fig. 17 shows students' learning gains on the conceptual knowledge posttest. Students performed significantly better on an immediate posttest assessing their conceptual knowledge about fractions compared to an equivalent pretest, scoring about 10% higher ($p < .01$, $d = .40$), as well as on a delayed posttest administered a week later, scoring about 15% higher ($p < .01$, $d = .60$). Fig. 18 shows students' learning gains on a test assessing procedural knowledge. With regard to procedural knowledge, students also performed significantly better on an immediate posttest compared to an equivalent pretest ($p < .01$, $d = .20$) and on a delayed posttest ($p < .01$, $d = .24$). Both the conceptual knowledge test and the procedural knowledge tests included transfer problems in which students had to apply their knowledge about fractions to novel task types.

Taken together, these findings show that Fractions Tutor is a successful intelligent tutoring system that yields significant and robust learning gains especially on conceptual knowledge tests that include transfer items. It is usable within real classroom settings and addresses the goals and needs of both students and teachers.

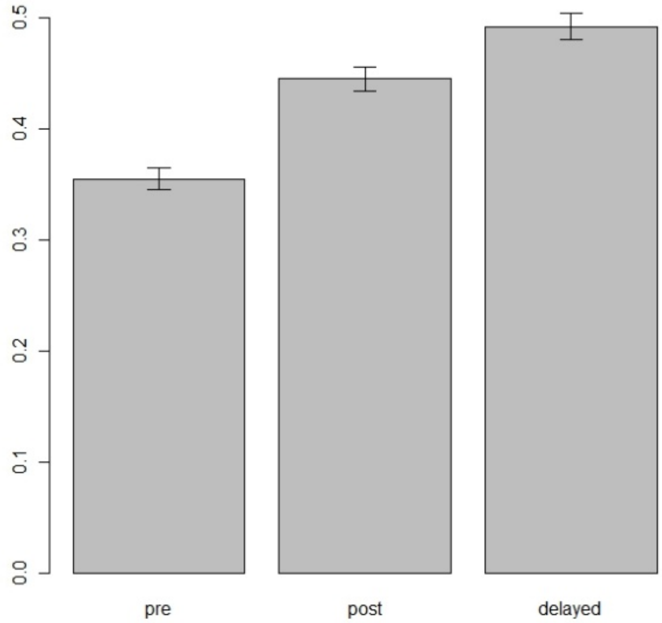


Fig. 17. Learning gains on the conceptual knowledge test.

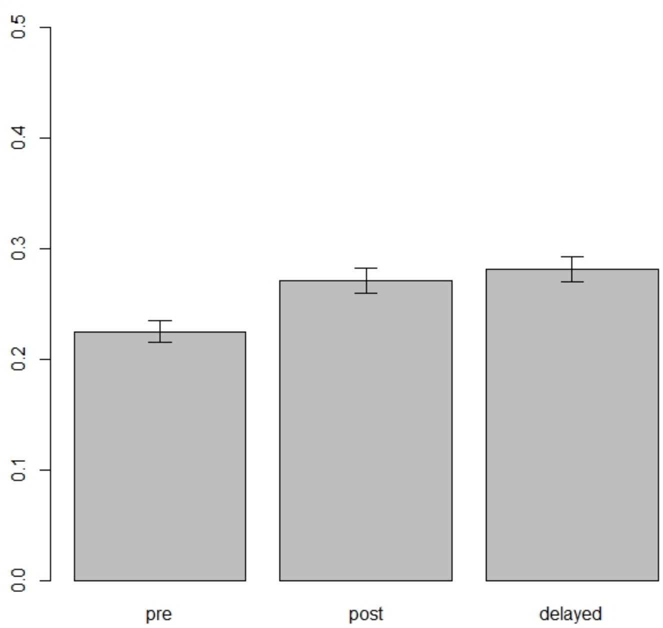


Fig. 18. Learning gains on the procedural knowledge test.

4 Classroom Experiments and Lab Studies

In this section, I describe a series of classroom experiments and lab studies that investigate (1) how best to support students' learning with multiple graphical representations, and (2) to iteratively improve the Fractions Tutor (see section 3). Furthermore, I reflect on the results from each experiment in the light of the theoretical framework on processes involved in learning with multiple graphical representations (see section 2).

Principle	Implementation	Experiment	Measure
Use multiple graphical representations to enhance robust learning of domain knowledge	Circles, rectangles, number lines	Experiments 1, 3, 4	Reproduction of conceptual fractions knowledge, transfer of procedural knowledge (Experiment 1); reproduction with number lines, transfer of conceptual knowledge (Experiment 3); conceptual knowledge (Experiment 4)
Use reflection prompts to support students in relating graphical representations to the corresponding symbolic representation	Menu-based reflection prompts	Experiment 1	Reproduction of conceptual fractions knowledge, transfer of procedural knowledge
Interleave task types while blocking graphical representations	Frequently switch between different task types	Experiment 2	Accuracy and efficiency of representational knowledge
Interleave graphical representations in addition to (moderately) interleaving task types	Frequently switch between graphical representations	Experiment 3	Conceptual transfer
Combine connectional sense-making support and connectional fluency-building support	Worked examples and mixed representations problems	Experiment 4	Conceptual knowledge
Implement connectional sense-making support in a way that requires students to actively generate connections	Worked examples	Experiment 4	Conceptual knowledge
Provide connectional sense-making support before connectional fluency-building support	Provide worked examples before mixed representations problems	Experiment 5	Accuracy of transfer of fractions knowledge

Table 3. Instructional design principles and their implementation in the Fractions Tutor.

Taken together, the experiments lead to a set of instructional design principles for the effective use of multiple graphical representations within intelligent tutoring systems. Each principle is based on experimental evidence for the effectiveness of specific implementations of instructional support for learning with multiple graphical representations. Table 3 gives an overview of the instructional design principles that follow from the experiments, how these principles are implemented in the Fractions Tutor, and on which measures of learning outcomes and learning processes these types of support were found to be effective in each experiment.

4.1 Experiment 1: Advantage of multiple graphical representations and self-explanation prompts

As discussed in section 1.1, research shows that multiple representations can significantly enhance students' learning: students typically learn better from a combination of text and graphical representation than from text alone (Ainsworth & Loizou, 2003; Baetge & Seufert, 2010; Bodemer et al., 2005; Butcher, 2006; Eilam & Poyas, 2008; Eitel et al., 2013; Kuehl et al., 2010; Magner et al., 2010; Mason et al., 2013; Rasch & Schnotz, 2009; Suthers et al., 2008). Furthermore, there is evidence that the positive effect of learning with multiple 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 multiple representations (Ainsworth & Loizou, 2003). Based on these findings, Ainsworth and Loizou (2003) hypothesize that multiple representations are beneficial because they can promote the self-explanation effect. Berthold and colleagues (Berthold et al., 2008; Berthold & Renkl, 2009) built on this work and investigated whether promoting self-explanation activities can enhance students' learning from multiple representations. They 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) and found that prompting promoted conceptual and procedural knowledge. Zhang and Linn (2011) evaluated an intervention that enhanced student-generated explanations while learning with dynamic chemistry representations. They found that the intervention helped students relate domain-relevant concepts to the visualizations. Taken together, this prior research on multiple representations indicates (1) that multiple representations can enhance students' learning by prompting reflection activities, and (2) that prompting students to reflect on relations between domain-relevant concepts and representations can further enhance students' benefits from multiple representations. However, neither of these studies systematically investigated whether the advantage of learning with multiple representations (compared to a single representation) can be enhanced by providing self-explanation prompts, or whether the advantage of multiple representations depends on students receiving such prompts.

Furthermore, all these prior studies used a symbol-systems approach: they included representations from *different* symbol systems, such as text and one additional graphical representation. As discussed in section 2, it remains an open question whether these findings generalize to learning with *multiple graphical* representations that use *the same* symbol system. Thus, the question of whether multiple graphical representations lead to better learning than a single graphical representation (provided in addition to symbolic and textual representations) remains open.

Experiment 1[1] addresses these questions and thereby lays the foundation for my dissertation research. Specifically, Experiment 1 systematically investigates whether the advantage of multiple representations generalizes to more complex, multi-representational learning materials, which are commonly used in real educational settings: *multiple graphical* representations. Furthermore, Experiment 1 tests whether students' benefit from multiple graphical representations can be enhanced by prompting them reflect on the relation between each graphical representation and the symbolic representation.

One-hundred thirty-two 6th-grade students worked with one of four versions of the Fractions Tutor 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 graphical representation versus multiple graphical representations) and reflection prompts (with versus without prompts). The reflection prompts were designed to encourage students to self-explain how the graphical representation relates to the symbolic notation while emphasizing knowledge components such as numerator and denominator. Students in the prompted conditions were asked to reflect on what aspects of the given graphical representations correspond to the numerator and the denominator of the fraction (e.g., "How does the number line show the numerator of the fraction?"), or how the procedure they performed symbolically corresponds to the manipulation of the graphical representations (e.g., "How did you convert the fraction in the circle?"). Students selected their answer from a drop-down menu, as shown in Fig. 19 – a procedure that has been shown to be effective many empirical studies with Cognitive Tutors (see Aleven & Koedinger, 2002; Koedinger, Aleven, Roll, & Baker, 2009). Students in the no-prompt conditions received the same tutor problems without the prompts. Students solved each problem by manipulating

[1] Experiment 1 was part of my diploma thesis (the German equivalent to a Master thesis), see Rau (2008).

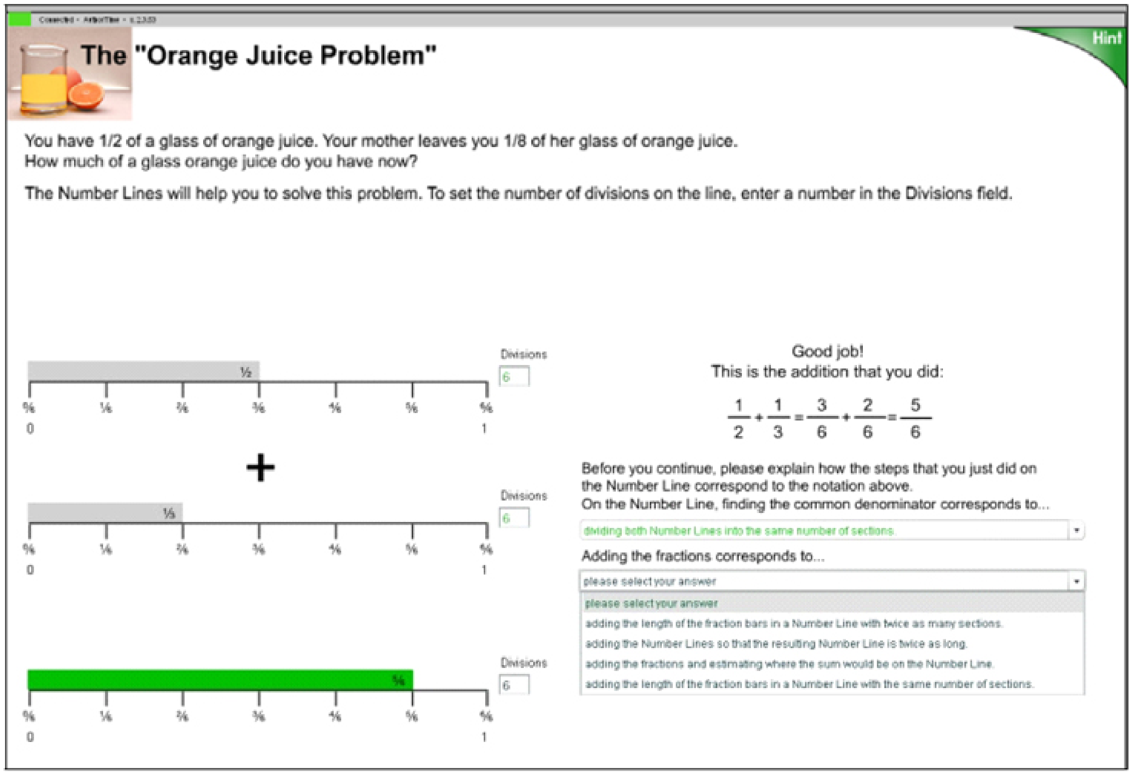


Fig. 19. Fraction addition with the number line representation, with self-explanation prompts.

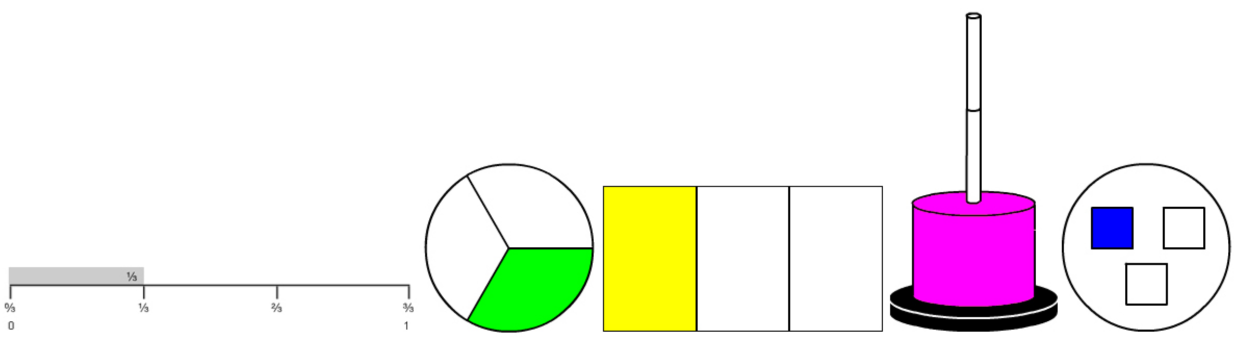


Fig. 20. Number line, circle, rectangle, stack, set (from left to right) as used in the Experiment 1.

both the symbolic representation of fractions, and a graphical representation.

In the single-representation conditions, all problems involved an interactive number line representation (see Fig. 20). In the multiple-representation conditions, students worked with five graphical representations: number lines, circles, rectangles, sets, and stacks (see Fig. 20). The different representations were presented in an interleaved fashion, so that only one graphical representation was presented at a time, but consecutive problems used different graphical representations. Students first solved a fractions problem using an interactive number line (see Fig. 19).

They then performed the same steps symbolically. Next, students revisited the same problem they had solved with the number line with the four remaining graphical representations.

Before working on the Fractions Tutor, students completed a prior knowledge test. The prior knowledge test took about 20 minutes to complete. On the following day, all students started working on the Fractions Tutor. Students worked on the equivalent fractions and fraction addition topics of the Fractions Tutor. Students worked with the Fractions Tutor for a total of 2.5 hours during two consecutive school days as part of their regular mathematics instruction. All students worked with the Fractions Tutor at their own pace, but the time they spent with the Fractions Tutor was held constant across experimental conditions, such that students completed as many tutor problems as they could in the time-frame available during the class periods. Immediately after finishing the work on the Fractions Tutor, students completed the immediate posttest, which took about 30 minutes. Six days after the immediate posttest, students completed the delayed posttest.

Results from 112 students show no main effect of number of graphical representations, but a main effect of reflection prompts on reproduction of conceptual knowledge of fractions as well as a significant interaction between number of graphical representations and reflection prompts on reproduction of conceptual knowledge, and on transfer of procedural knowledge at the immediate posttest. Specifically, students in the prompted conditions performed better if they had worked with multiple graphical representations than if they had worked with a single graphical representation, whereas students within the no-prompt conditions performed worse if they had worked with multiple graphical representations than if they had worked with a single graphical representation. Please refer to Rau and colleagues (2009) or Rau (2008) for a more detailed description of Experiment 1.

Taken together, the Experiment 1 extends prior research on multiple representations that use *different* symbol systems and shows that multiple graphical representations that use *the same* symbol system can enhance students' robust learning of domain knowledge, provided that students are prompted to self-explain the relation between graphical and symbolic representations. This finding extends the prior theoretical frameworks for learning with multiple representations described in section 2.1. Since multiple graphical representations use the same symbol system, they do not require students to integrate information from a larger number of symbol systems than when students are provided with only a single graphical representation. Furthermore, the different graphical representations are all encoded into a verbal model within working memory and thus do not alter the cognitive capacity available. By consequence, Experiment 1 demonstrates that the advantage of learning with multiple representations is not limited to representations of different symbol systems or the number of information channels used to process the representations.

I attribute this finding to the complementary conceptual perspectives that multiple graphical representations provide on the learning content. As many STEM domains, fractions instruction uses different graphical representations with the goal to emphasize different conceptual aspects of the domain (e.g., fractions as measurements in number lines and fractions as parts of a whole in area models; Charalambous & Pitta-Pantazi, 2007). In order to understand the concept of fractions, students need to integrate the conceptual views depicted by the different graphical representations. This notion (i.e., that deep conceptual processing of the structural elements that constitute the complex learning material is crucial to students' benefit from multiple graphical representations) does not contradict Schnotz and Bannert's (2003) framework. Rather, it extends Schnotz and Bannert's (2003) work by demonstrating that this type of integration process does not have to occur across *different* symbol systems: it can also occur between multiple representations that are part of *the same* symbol system.

However, conceptual integration does not (often) happen spontaneously (Ainworth, 2006; Yerushamly, 1991). Experiment 1 systematically investigates whether reflection prompts designed to encourage students to relate graphical to symbolic representations of fractions enhance their benefit from multiple graphical representations. The results show that such prompts are *necessary* for students to benefit from multiple graphical representations: only when provided with self-explanation prompts did I find an advantage of multiple over a single graphical representation. Experiment 1 is, to the best of my knowledge, the first to systematically investigate the effects of self-explanation prompts and multiple versus single graphical representations.

The findings from Experiment 1 are also of practical importance: they provide guidelines for developers of instructional materials (such as intelligent tutoring systems) that include multiple graphical representations. Students will benefit from multiple graphical representations *only* if they are prompted to relate graphical and symbolic representations. Reflection prompts are an appropriate means (though perhaps not the only effective one) to support students in doing so. As a consequence, the Fractions Tutor includes reflection prompts, as described in section 3.4.1.

With regard to the processes involved in learning with multiple graphical representations (see section 2.2), the process supported by the reflection prompts may most likely be characterized as a representational sense-making process. The reflection prompts used in Experiment 1 encourage students to explicitly relate the graphical representations to the symbolic representation based on the abstract concept they both depict. By mapping abstract concepts (e.g., the numerator of a fraction) to the corresponding part of the symbolic notation of fractions (i.e., the top number) and to the corresponding component of the graphical representation (e.g., the number of shaded pieces in a circle), students are supported in acquiring representation-specific knowledge components (e.g., ident-numerator-circle). Thus, one might argue, based on the results in Experiment 1, that support for representational sense-making processes (e.g., in the form of reflection prompts) enhance students' learning of robust domain knowledge with multiple graphical representations. However, Experiment 1 only assessed students' domain knowledge as learning *outcomes*, but not their representational understanding during the learning *process*. To truly put this interpretation (i.e., that support for representational sense-making processes enhance students' benefit from multiple graphical representations) to a test, one would also have to assess students' representational understanding. One might expect that reflection prompts enhance representational understanding, and further, that differences in representational understanding between groups with versus without such prompts explain the differences in their performance on domain knowledge tests.

In conclusion, Experiment 1 extends prior research on multiple representations that has been conducted under the symbol-systems approach to the case of multiple graphical representations, which all use the same symbol system. By demonstrating that multiple graphical representations can enhance learning (when accompanied by reflection prompts) provides the foundation for the experimental studies described in the remainder of this section.

4.2 Experiment 2: Interleaving task types while blocking graphical representations

In multi-representational educational technologies, learners typically engage in extended problem-solving practice with multiple graphical representations across several task types. In these cases, instructors and instructional designers must decide how to sequence graphical representations (e.g., circles, number lines) and task types (e.g., finding equivalent fractions and comparing fractions). Should they interleave multiple graphical representations while blocking task types, or should they interleave task types while blocking multiple graphical representations? What sequence will lead to the most robust learning gains? The decision of how to sequence task types and graphical representations is likely to influence learners' acquisition of robust conceptual and procedural knowledge.

The question of whether to interleave graphical representations or task types is not only of practical importance. Although the advantages of learning with multiple representations are well-documented, this research has not yet investigated the effects of interleaved practice with multiple representations. The literature on contextual interference has demonstrated that the temporal sequence of learning tasks affects students' robust learning (Battig, 1972; Schmit & Bjork, 1992): interleaving different learning tasks (rather than blocking them) leads to better long-term retention and better performance on transfer tests. However, this research has not yet investigated whether the dimension on which the learning tasks are interleaved (e.g., task type or graphical representation) matters. The question of which dimension instructional designers should interleave is therefore of both practical and theoretical importance.

4.2.1 Research questions and hypotheses

The literature on contextual interference suggests that the decision of whether to interleave multiple graphical representations or task types will influence students' robust learning of domain knowledge. Generally, results from contextual interference research show that interleaved practice leads to better learning outcomes than blocked practice (Battig, 1972; Schmidt & Bjork, 1992). In this research, the independent variable typically is whether learning tasks are 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 – task 3), or whether learning tasks of different types are 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 can be found in a variety of domains including vocabulary learning (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Pashler, Rohrer, Cepeda, & Carpenter, 2007), motor tasks (Hebert, Landin, & Solmon, 1996; Immink & Wright, 1998; Li & Wright, 2000; Meiran, 1996; Meiran, Chorev, & Sapir, 2000; Ollis, Button, & Fairweather, 2005; Schmidt & Bjork, 1992; Shea & Morgan, 1979; Simon & Bjork, 2001), algebra (Rohrer, 2008; Rohrer & Taylor, 2007; Taylor & Rohrer, 2010), troubleshooting (de Croock, van Merriënboer, & Paas, 1998; van Merriënboer, Schuurman, de Croock, & Paas, 2002), and decision-making tasks (Helsdingen, van Gog, & van Merriënboer, 2011). However, the research on interleaved practice has not investigated whether the dimension on which learning tasks are interleaved (i.e., interleaving graphical representations versus interleaving task types) matters. In other words, it remains an open question whether multiple graphical representations are more effective when they (the different graphical representations) are interleaved or when task types are interleaved.

In particular, the advantage of interleaved practice has been attributed to two kinds of processes that play a role in deep, cognitive processing of the learning material (Rau, Aleven et al., 2013b). First, interleaved practice schedules require learners to frequently reactivate the knowledge needed to solve each learning task (de Croock et al., 1998; Lee & Magill, 1983, 1985): when tasks are presented in an interleaved sequence, the required knowledge has to be retrieved more frequently from long-term memory. Retrieval from long-term memory strengthens the association between cues and associated elements in long-term memory, and increases the likelihood that this knowledge can be recalled later on (Anderson, 1993; Anderson, 2002). Second, interleaving may help students abstract knowledge across different learning tasks (de Croock et al., 1998; Shea & Morgan, 1979). When knowledge needed for different learning tasks is co-active in working memory, students can compare the knowledge relevant to the respective learning tasks. While this process may happen consciously or unconsciously, it helps learners to see which task properties are key and which are incidental, thereby directing their attention to aspects relevant to knowledge construction (Bannert, 2002; Paas & van Gog, 2006; van Merriënboer et al., 2002).

How, based on these theoretical accounts, might one expect interleaving task types or graphical representations to affect students' robust learning? I hypothesize that it will be most effective to interleave learning tasks along the dimension of greatest variability, that is, the dimension along which the learning tasks vary the most from task to task. For two reasons, I hypothesize that task types (e.g., equivalent fractions, fraction comparison, fraction addition, etc.) are more variable than the graphical representations used in the Fractions Tutor. First, the different task types require students to apply different operations (e.g., finding equivalent fractions, adding fractions). By contrast, the different graphical representations provide different conceptual views on the task at hand (by depicting a fraction as a shaded part of a circle, or as a dot on the number line), and these conceptual differences might be difficult for novice learners to discern. Second, the graphical representations are designed to be intuitive: graphical representations typically employ perceptual processes in an easy-to-understand way. They may also be intuitive in the sense that they connect to students' informal prior knowledge about fractions. To use graphical representations, students are not expected to engage in explicit reasoning about the properties of the representations. The different task types, by contrast, require students to explicitly use different procedures to solve the task. Due to these properties of graphical representations and task types, I expect that the conceptual differences between graphical representations will not be as salient as the differences between task types. For this reason, I anticipate that task types are the more variable dimension, compared to graphical representations.

Interleaving learning tasks along the more variable dimension should give students better opportunities for reactivation and abstraction. Students are expected to reactivate any knowledge that is not shared between consecutive learning tasks. I expect this reactivation process to happen more frequently when learning tasks differ on the more variable dimension. Since repeated reactivation of knowledge strengthens that knowledge and increases the chance that it can be recalled later on, reactivation is, in turn, expected to increase students' acquisition of robust knowledge.

Abstraction may be more likely to occur when learning tasks are interleaved on a moderately variable dimension. Consecutive tasks need to be sufficiently dissimilar so that students can compare the knowledge associated with the different learning tasks: without dissimilarities, there is nothing to abstract across. However, it is crucial for abstraction to occur that the consecutive learning tasks share some common knowledge that can be abstracted from them. If the learning tasks are too dissimilar, students may not be able to abstract common knowledge from them. In other words, if abstraction is the mechanism by which interleaved practice leads to better learning, higher variability of consecutive learning tasks should not always lead to better learning outcomes. Rather, there might be an optimal level of variability in learning tasks such that consecutive tasks are similar enough to allow for abstraction, but dissimilar enough so that learners are likely to abstract knowledge from them. Since the different learning tasks employed in the Fractions Tutor share a substantial amount of knowledge between them (by virtue of covering conceptual and procedural aspects of fractions), I do not expect that the variability of these learning tasks is too high to prevent abstraction.

Given that these arguments are not specific to conceptual or procedural knowledge, I expect that interleaving task types (while blocking graphical representations) will have a stronger effect on students' acquisition of conceptual and of procedural knowledge than interleaving graphical representations (while blocking task types).

4.2.2 Methods

To investigate these questions, I conducted a classroom experiment that contrasted different practice schedules of graphical representations and task types.

4.2.2.1. Experimental design

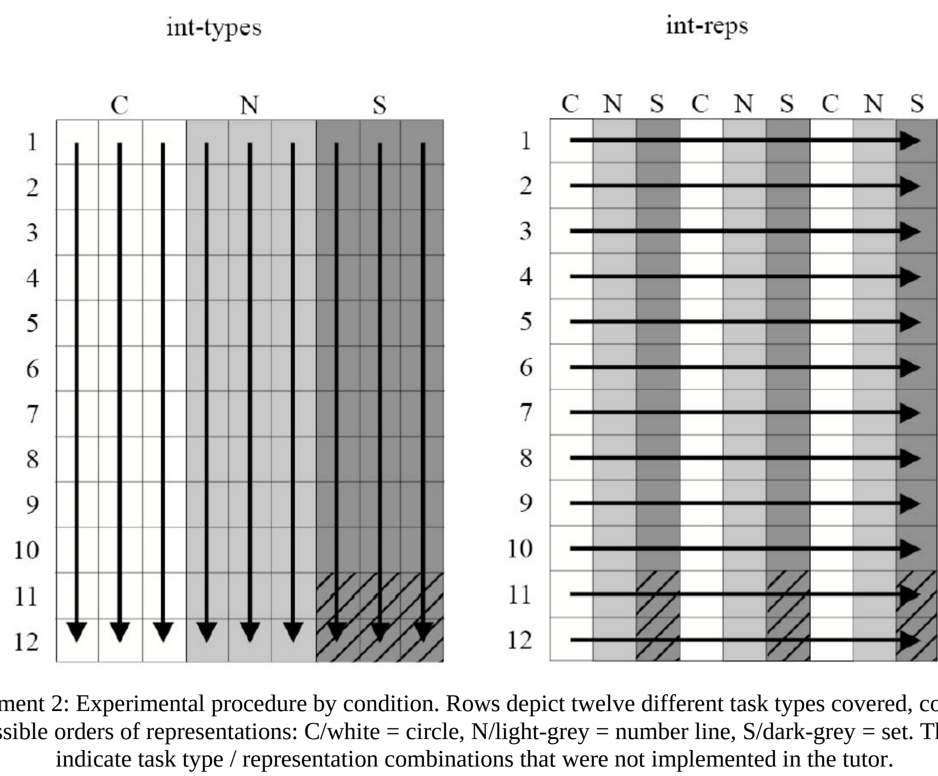


Fig. 21. Experiment 2: Experimental procedure by condition. Rows depict twelve different task types covered, columns show one of several possible orders of representations: C/white = circle, N/light-grey = number line, S/dark-grey = set. The dashed cells indicate task type / representation combinations that were not implemented in the tutor.

The goal of this study was to systematically investigate the effects of interleaving task types (int-types) versus interleaving representations (int-reps). Students were randomly assigned to one of two conditions. In the *int-types* condition, the task types were interleaved while the graphical representations were blocked. In the *int-reps* condition, the graphical representations were interleaved while the task types were blocked. Students in all conditions worked on the same 102 fractions tasks at their own pace, with the help from the Fractions Tutor. All learning tasks involved one individual graphical representation per tutor problem, but multiple across a sequence of tutor problems (as described in section 3.4.1). Each problem also involved the symbolic representation of fractions and a problem statement in text. Fig. 21 clarifies how the conditions were implemented. Each table represents the set of 102 problems that students solved with the tutor. Each row represents one of twelve task types (e.g., equivalent fractions, or fraction addition). There were nine problems for each task type (i.e., each row stands for nine problems). Each representation was coupled with each task type – there were three problems for any such combination[2]. Thus, the number of problems of each type, the number of problems with each representation, and the number of problems that couple a particular task type and representation are constant across conditions. In the int-types condition (see the table on the left in Fig. 21), the task types are maximally interleaved and the representations are maximally blocked. That is, students covered all twelve fraction task types with one graphical representation before switching to the next representation, again working through all task types before switching to the third graphical representation (corresponding to 36 problems per representation). In this condition, students encountered a new task type after every single problem. By contrast, in the int-reps condition (see the table on the right in Fig. 21), the representations were maximally interleaved and the task types were maximally blocked. That is, students worked on all problems of one task type (covering it with all three representations) before moving on to the next task types. In this condition, students encountered a different graphical representation after every single problem. Thus, the degree of interleaving is the same across conditions; what varies is what is being interleaved.

[2] All task types were presented with each graphical representation with the exception of two fraction addition task types where the use of the set representation is not advisable from an instructional standpoint. The exclusion of the set representation from two of twelve task types does not change the level of blocking or interleaving of task types or representations and therefore does not interfere with the intervention.

In order to prevent possible order effects, I implemented different plausible orders of graphical representations as a control factor to counterbalance potential ordering effects. Students never worked with the set representation first because sets appear to be the graphical representation with which students are least familiar (i.e., presenting students with the set representation first cannot be recommended from an instructional perspective and thus does not represent a realistic educational scenario). Students were randomly assigned to one of four different orders of graphical representations: circle – number line – sets, circle – set – number line, number line – circle – set, or number line – set – circle. Fig. 21 thus reflects only one of the implemented orders

4.2.2.2. Participants

The study involved 158 students in grades 5 and 6, aged 9 to 12 years, from 16 classes of a total of three schools. Students participated in the study during their regular math instruction.

4.2.2.3. Procedure

Experiment 2 took place at the end of the school year 2008/2009. Students' regular math teachers led the sessions, but researchers were present in the classrooms at all times to assist teachers in answering questions specific to the use of the tutoring system.

Students' knowledge of fractions was assessed three times. On the first day, students completed a pretest. They then worked on twelve task types taken from six topics of the Fractions Tutor (with the version depicted in Fig. 9, see section 3.3.3.3; topics 1, 2, and 6-9 in Table 1, see section 3.4), for five hours, spread across five to six (depending on specific school schedules) consecutive days. The day following the tutor sessions, students completed an immediate posttest. Seven days later, in order to assess whether students' learning is robust in that it lasts over time (see Koedinger et al., 2012), students completed an equivalent delayed posttest. Students could take as much time as they needed to complete the tests. Participating teachers were asked not to revisit fractions between the immediate and the delayed posttest.

4.2.2.4. Test instruments

To assess students' robust knowledge of fractions, I created a test that included two scales: *representational knowledge* and *operational knowledge*, described further below. The theoretical structure of these tests (i.e., the division of the test items into representational and operational knowledge) was validated by a confirmatory factor analysis using data from a large sample of students collected during a pilot study (i.e., a different sample than participated in the current study).

Each of the two test scales included both familiar and unfamiliar tasks (i.e., task types that students had encountered during their work on the tutor and task types that were new relative to those covered in the tutor). The goal in including the latter types of tasks was to assess whether students acquired robust knowledge that can be transferred to unfamiliar problems (see Koedinger et al., 2012). Appendix 5 shows a sample test item for the representational knowledge and the operational knowledge scales, respectively. The representational knowledge scale of the test assessed students' conceptual knowledge of fractions representations. I operationalized representational knowledge as the ability to interpret representations in terms of fractions, including graphical representations that were not covered by the tutor. All items of the representational knowledge test scale included graphical representations, including representations that students did not encounter in the set of tutor problems: fraction strips, and contextualized applications of measurement scales, analog clocks, and concrete objects. By contrast, the operational knowledge scale assessed students' procedural knowledge of fractions operations. I operationalized operational knowledge as students' ability to perform familiar operations (i.e., operations they had practiced in the tutor, such as fraction addition) either without graphical representations or with an unfamiliar graphical representation (i.e., fraction strips). The operational items also included items that required operations that were not covered by the tutor (i.e., fraction subtraction) solved without graphical representations. Both test scales included items that I adapted from standardized tests in the United States (NAEP, PSSA) and from examples from the fractions literature (Rittle-Johnson & Koedinger, 2005).

Two different equivalent versions (version A and version B) of the test were created. The test versions included the same tasks but used different numbers. The pilot study of the test instruments showed that the test versions did significantly not differ in terms of difficulty. I randomly assigned students to either version A or B of the fractions test at the pretest, assigned them the other version at the immediate posttest, and randomly assigned either version A or B at the delayed posttest.

I assessed students' robust knowledge of fractions using both accuracy and efficiency measures. The accuracy measure corresponded to the mean score on the representational knowledge scale of the test and the operational knowledge scale of the test, respectively. To analyze students' efficiency on the tests, I used a measure of efficiency described by (van Gog & Paas, 2008) and by (Lewis & Barron, 2009). Specifically, I combined students' standardized average scores on the representational knowledge and the operational knowledge subscales of the test and the standardized average time they spent on each of the test subscales across pretest, posttest, and delayed posttest using the following formula:

$$\text{efficiency (subscale of test)} = \frac{\text{Z (score on subscale of test)} - \text{Z (time spent on subscale)}}{\sqrt{2}} \quad (1)$$

Positive efficiency scores indicate higher efficiency at solving quiz items correctly, and negative efficiency scores indicate lower efficiency at solving quiz items correctly, compared to the relative mean of the sample. An efficiency score of 0 indicates average efficiency with respect to the sample.

I followed van Gog and Paas (2008) and Lewis and Barron (2009) and applied the concept of condition efficiency (Paas & van Merriënboer, 1993) to a measure of performance efficiency. Paas and van Merriënboer (1993) used performance and mental effort to compute efficiency. Van Gog and Paas (2008) argue that time on task can also be viewed as an approximation of mental effort. I used the time students spent on the test rather than the time they spent with the tutoring system for two reasons. First, I was interested in students' efficiency in answering test items, rather than in how efficiently they learn, because the ability to solve a test fast and accurately is required in many assessment situations, for example in standardized tests in the United States. Second, using time spent on the tutoring system as the measure of mental effort during the learning phase depends on the assumption that time-on-task during the learning phase was not restricted. This assumption does not hold, however, because the students worked with the tutoring system during their regular math periods, which are, due to their nature, restricted in time.

4.2.3 Results

Table 4 provides the means and standard deviations for the accuracy of representational and operational knowledge, for time-on-task on the representational and operational knowledge subscales of the test, and for representational efficiency and operational efficiency.

		pretest	immediate posttest	delayed posttest
representational accuracy	int-types	.55 (.20)	.61 (.27)	.59 (.24)
	int-reps	.53 (.22)	.52 (.26)	.39 (.32)
operational accuracy	int- types	.38 (.31)	.51 (.34)	.44 (.36)
	int-reps	.43 (.33)	.40 (.35)	.39 (.30)
representational time-on-task	int-types	100.18 (29.53)	70.89 (22.46)	66.95 (23.60)
	int-reps	103.63 (32.19)	74.88 (28.33)	79.24 (30.52)
operational time-on-task	int- types	68.74 (31.64)	47.69 (18.53)	43.88 (19.09)
	int-reps	69.77 (37.01)	46.11 (21.31)	42.71 (22.22)
representational efficiency	int-types	-.35 (.89)	.48 (.83)	.52 (.80)
	int-reps	-.48 (.98)	.16 (.90)	-.31 (1.01)
operational efficiency	int- types	-.50 (.90)	.31 (.83)	.24 (.77)
	int-reps	-.42 (1.12)	.12 (.78)	.17 (.63)

Table 4. Means and standard deviations (in parentheses) for accuracy, time-on-task (in seconds), and efficiency for the representational and operational knowledge subscales at pretest, immediate posttest, delayed posttest by condition in Experiment 2.

4.2.3.1. Analysis

Students were excluded if they were not present on all test days (n = 49), if they worked on the tutoring system during the weekend (n = 1), if they had an overall pretest score of 0.95 or higher (n = 2), or if I did not have information on how much time they spent on each item of the test (n = 5). Table 5 shows the number of included and excluded students per condition.

	included	excluded
int-types	52	27
int-reps	49	30

Table 5. Experiment 2: Number of students included and excluded by condition.

After excluding these students, a total of N = 101 remained in the sample (n = 52 for int-types and n = 49 for int-reps). The number of excluded students did not differ between experimental conditions, χ^2 (1, N = 158) < 1, nor did the time spent on the tutor problems (F < 1). A MANOVA on the pretest scores showed that students who were excluded from the analysis scored significantly higher on the representational knowledge scale of the test, $F(1, 156) = 13.192$, $p < .01$, and on the operational knowledge scale of the test, $F(1, 156) = 6.456$, $p < .05$, than students who were included in the analysis. No significant differences between conditions were found at the pretest for representational knowledge (F < 1), or operational knowledge (F < 1). Since there was no effect for order of representation on representational knowledge (F < 1) or operational knowledge $F(3, 97) = 1.21$, $p > .10$, I disregarded the order of representation in the following analyses. Since students had seen the same test that they received at the delayed posttest either at the pretest or at the immediate posttest, I analyzed the effect of having seen the same test form either at the pretest or at the immediate posttest. There was no significant difference between students for the time (i.e., either at the pretest or at the immediate posttest) students had seen the same test before on representational knowledge ($F < 1$) or operational knowledge ($F < 1$). Finally, because some students did not finish all problems on the tutor in the time given, I computed a covariate that describes, for each student, the number of tutor problems solved that involved the knowledge components tested by the representational and by the operational knowledge tests, respectively.

A hierarchical linear model (HLM; Raudenbush & Bryk, 2002) with four nested levels was used to analyze the data. At level 1, I modeled performance for each of the two posttests for each student. At level 2, I accounted for differences between students. At level 3, I modeled random differences between classes, and at level 4, I accounted for random differences between schools. The HLM is the outcome of a forwards-inclusion procedure in which I used the Bayesian Information Criterion (BIC) to find whether the inclusion of a variable increased model fit. A lower BIC indicates better model fit while penalizing for greater model complexity. If the BIC decreased as a consequence of including a variable, I kept the variable. If the BIC did not decrease, I did not include the variable. I tested a number of variables, including teacher, sequence of graphical representations, test form sequence, grade level, number of problems completed, time per tutor step, interaction of test time with condition, interaction of pretest with condition, random intercepts and slopes for classes and schools. Equation 2 describes the resulting HLM that I fitted to the data:

$$Y_{ijkl} = (((\mu + W_l) + V_{kl}) + \beta_3 * c_j + \beta_4 * p_j + \beta_5 * e_j + U_{jkl}) + \beta_1 * t_i + \beta_2 * c_j * t_i + R_{ijkl} \qquad (2)$$

with

(level 1) $Y_{ijkl} = \varepsilon_{jkl} + \beta_1 * t_i + \beta_2 * c_j * t_i + R_{ijkl}$

(level 2) $\varepsilon_{jkl} = \delta_{kl} + \beta_3 * c_j + \beta_4 * p_j + \beta_5 * e_j + U_{jkl}$

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

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

with the index *i* standing 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$_j$'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 the interaction of condition c_j with test time t_i, β_3 is the parameter for the effect of condition c_j, β_4 is the parameter for the effect of student$_j$'s performance on the pretest p_j, β_5 is the parameter for the effect of exposure e_j, which indicates how many problems student$_j$ solved on the relevant knowledge components being tested by either the representational knowledge test or the operational knowledge, 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.

In addition, I specified posthoc comparisons within the HLM to clarify the effects of condition. All reported *p*-values were adjusted using the Bonferroni correction. I report partial η^2 for effect sizes on effects including more than two conditions, and Cohen's *d* for effect sizes of pairwise comparisons. According to (Cohen, 1988), an effect size partial η^2 of .01 corresponds to a small effect, .06 to a medium effect, and .14 to a large effect. An effect size *d* of .20 corresponds to a small effect, .50 to a medium effect, and .80 to a large effect.

4.2.3.2. Effects of practice schedules

measure	**test time**	**main effects / interaction effects**	**tendency of pairwise comparisons**	**significant (yes/no)**	***F*/*t*-value**	**adj. *p*-value**	**effect size**
repre-sentational accuracy		condition		yes	$F(1, 100) = 23.97$	$p < .01$	partial η^2 = .11
		test time		yes	$F(1, 100) = 4.99$	$p < .05$	partial η^2 = .01
		condition * test time		yes	$F(1, 100) = 7.32$	$p < .01$	partial η^2 = .05
	immediate posttest		int- types > int-reps	yes	$t(100) = 2.34$	$p < .05$	$d = .37$
	delayed posttest		int- types > int-reps	yes	$t(100) = 5.55$	$p < .01$	$d = .88$
repre-sentational efficiency		condition		yes	$F(1, 100) = 18.28$	$p < .01$	partial η^2 = .07
		test time		yes	$F(1, 100) = 4.94$	$p < .05$	partial η^2 = .02
		condition * test time		yes	$F(1, 100) = 7.46$	$p < .01$	partial η^2 = .03
	immediate posttest		int- types > int-reps	yes	$t(100) = 2.03$	$p < .05$	$d = .09$
	delayed posttest		int- types > int-reps	yes	$t(100) = 4.74$	$p < .01$	$d = .21$

Table 6. Experiment 2: Results on differences between conditions on representational accuracy and representational efficiency.

To investigate the effect of practice schedules on accuracy of representational knowledge, I applied the HLM in equation 2 to students' accuracy scores on the representational knowledge subscale of the test. Table 6 provides an overview of the learning results on both the representational accuracy and representational efficiency measures. Table 7 summarizes the least squared means and standard deviations generated by the HLM for representational accuracy and representational efficiency. There was a significant main effect for condition on representational accuracy, $F(1, 100) = 18.28$, $p < .01$, partial $\eta^2 = .07$. There was also a significant main effect of test time (i.e., immediate posttest and delayed posttest), $F(1, 100) = 7.46$, $p < .05$, partial $\eta^2 = .02$. The main effects were qualified by a significant interaction between test time (i.e., immediate or delayed posttest) and condition, $F(1, 100) = 4.94$, $p < .01$, partial $\eta^2 < .03$. To gain insights into the nature of this interaction, I computed posthoc comparisons for the effect of condition at the immediate posttest and the delayed posttest, respectively. On representational accuracy, there was an advantage for int-types over int-reps on the immediate posttest, $t(100) = 2.03$, $p < .05$, $d = .09$, and the delayed posttest, $t(100) = 4.74$, $p < .01$, $d = .21$. Taken together, these results show that the int-types condition outperforms the int-reps condition on accuracy of representational knowledge.

		immediate posttest	**delayed posttest**
representational accuracy	int-types	.58 (.05)	.57 (.05)
	int-reps	.50 (.05)	.36 (.05)
representational efficiency	int- types	.41 (.22)	.45 (.22)
	int-reps	.03 (.22)	-.43 (.22)

Table 7. Experiment 2: Least squared means and standard deviations (in parentheses) for representational accuracy and representational efficiency at immediate posttest, delayed posttest by condition.

To investigate the effect of practice schedules on efficiency of representational knowledge, I applied the HLM in equation 2 to students' efficiency scores on the representational knowledge subscale of the test. I found a significant main effect for condition on representational efficiency, $F(1, 100) = 23.97$, $p < .01$, partial $\eta^2 = .11$. There was also a significant main effect of test time, $F(1, 100) = 4.99$, $p < .05$, partial $\eta^2 = .01$. The main effects were qualified by a significant interaction between test time (i.e., immediate or delayed posttest) and condition, $F(1, 100) = 7.32$, $p < .01$, partial $\eta^2 = .05$, such that the difference between conditions was stronger on the delayed posttest than at the immediate posttest. Posthoc comparisons between groups were computed to clarify the interaction effect at the immediate posttest and the delayed posttest, respectively. On representational efficiency, there was an advantage for int-types over int-reps on the immediate posttest, $t(100) = 2.34$, $p < .05$, $d = .37$, and the delayed posttest, $t(100) = 5.55$, $p < .01$, $d = .88$.

These findings show that the int-types condition outperforms the int-reps condition on efficiency of representational knowledge.

test scale	test time	main effects / interaction effects	tendency of pairwise comparisons	significant (yes/no)	*F*/*t*-value	adj. *p*-value	effect size
operational accuracy		condition		no	$F(1,100) = 2.05$	$p > .10$	
		test time		yes	$F(1,100) = 3.04$	$p < .10$	partial $\eta^2 < .01$
		condition * test time		no	$F(1,100) = 1.36$	$p > .10$	
	immediate posttest		int- types > int-reps	no	t < 1		
	delayed posttest		int- types > int-reps	no	$t < 1$		
operational efficiency		condition		no	$F < 1$		
		test time		no	F < 1		
		condition * test time		no	F < 1		
	immediate posttest		int- types > int-reps	no	t < 1		
	delayed posttest		int- types > int-reps	no	$t < 1$		

Table 8. Experiment 2: Results on differences between conditions, obtained from HLM described in equation 2.

		immediate posttest	delayed posttest
operational accuracy	int-types	.51 (.04)	.43 (.04)
	int-reps	.41 (.04)	.39 (.04)
operational efficiency	int- types	.25 (.11)	.18 (.11)
	int-reps	.17 (.10)	.23 (.10)

Table 9. Experiment 2: Least squared means and standard deviations (in parentheses) for operational accuracy and operational efficiency at immediate posttest, delayed posttest by condition.

To investigate the effects of practice schedules on operational knowledge, I applied the HLM in equation 2 to students' accuracy and efficiency scores on the operational knowledge subscale of the test. Table 8 provides an overview of the learning results on both the operational accuracy and operational efficiency measures. Table 9 summarizes the least squared means and standard deviations generated by the HLM for operational accuracy and operational efficiency. On operational accuracy, I found no significant main effect of condition, $F(1,100) = 2.05$, $p > .10$. The effect of test time was marginally significant for operational accuracy, $F(1,100) = 3.04$, $p < .10$, partial $\eta^2 < .01$. There was no significant interaction effect for operational accuracy, $F(1,100) =$

1.36, $p > .10$. These results show that the int-types condition does not outperform the int-reps condition on accuracy of operational knowledge.

To investigate the effect of practice schedules on *efficiency of operational knowledge*, I applied the HLM in equation 2 to students' efficiency scores on the operational knowledge subscale of the test. On operational efficiency, there was no significant main effect of condition, $F < 1$. The effect of test time was not significant for operational efficiency, $F < 1$. There was no significant interaction effect for operational efficiency, $F < 1$. Taken together, these findings show that the int-types condition does not outperform the int-reps condition on efficiency of operational knowledge.

Please also refer to Rau, Aleven, and colleagues (2013b) for the results from Experiment 2.

4.2.4 Discussion

As hypothesized, the results show that interleaving task types while blocking graphical representations leads to both higher accuracy and higher efficiency in answering questions that require representational knowledge, compared to interleaving graphical representations while blocking task types. Thus, overall, the results provide support for the notion that interleaving task types leads to more robust representational knowledge than interleaving graphical representations. On the other hand, the practice schedule of task types and graphical representations did not significantly affect students' robust learning of operational knowledge.

How might one explain the differences between conditions on representational knowledge? I argued that interleaving learning tasks along the dimension of greatest variability is most effective, and task types are the more variable dimension (compared to graphical representations) because the differences among task types are more salient than the differences between graphical representations. Although different graphical representations emphasize conceptual different aspects of fractions, students might not readily perceive these dissimilarities because the representations are designed to be intuitive and easy to interpret. Therefore, it might be difficult for novice students to discern the conceptual differences between the different graphical representations. Greater problem variability may also increase the number of opportunities for abstraction. If tasks are very similar, it may be difficult for students to abstract across them. It is possible that the subtle differences between graphical representations makes it difficult for students to abstract across them. Instead, interleaving task types may encourage students to abstract across different applications of the same graphical representations, which appears to lead to a most robust conceptual understanding of graphical representations than abstracting across different graphical representation. Applying the same representation to different subsequent task types may allow students to form an abstract understanding of the given representation independent of its application to a specific task type. Consequently, interleaving task types may have a larger impact on acquiring representational knowledge than interleaving representations. Taken together, Experiment 2 suggests that reactivation and abstraction across task types is more beneficial to students' conceptual understanding of graphical representations than abstracting across representations. Note that these two suggested mechanisms are not mutually exclusive, and both might account for the advantage of interleaving task types. Yet, Experiment 2 did not assess whether these mechanisms occur. Collecting think-aloud protocols while students work with the Fractions Tutor might shed light into the question of whether interleaving task types (compared to interleaving graphical representations) differ in the degree to which they enhance reactivation and abstraction processes.

The fact that interleaving task types versus interleaving graphical representations affects only representational knowledge but not operational knowledge may reflect differences in these knowledge types. The representational knowledge scale requires primarily conceptual knowledge about how to interpret representations, and the ability to apply this knowledge to new representations of the same underlying domain concepts. The operational knowledge scale assesses students' ability to apply procedures to solve fractions problems, and to transfer these procedures by adapting them to novel problems (including problems without representations). The results from Experiment 2 appear to indicate that practice schedules have a greater impact on conceptual understanding of graphical representations than on operational knowledge. Some other studies have failed to find effects of practice schedules altogether (e.g., French, Rink, & Werner, 1990; Jones & French, 2007). They argued that the effect of practice schedules depends on the complexity of the learning tasks because the complexity of the task impacts the processing demands (Shea & Morgan, 1979; Wulf & Shea, 2002). Although these studies have been conducted in a radically different domain, their argument may apply to Experiment 2 as well and may help to explain the lack of differences between conditions on operational knowledge. For instance, Wulf and Shea (2002) suggest that the higher the complexity of the learning tasks (which increases processing demands compared to low-complexity tasks), the less students will benefit from interleaved practice (which corresponds to a further increase of the processing demands). In Experiment 2, it may be that the operational knowledge covered by the tutor was more complex than the representational knowledge. Fractions operations, when carried out with graphical representations, may rely on an at least basic understanding of how the graphical representations depict fractions, in addition to knowledge about fractions operations. In other words, the operational knowledge may have required some representational knowledge, whereas the representational knowledge covered by the Fractions Tutor may not have required operational knowledge. It is therefore possible that the operational knowledge acquisition as supported by the Fractions Tutor was, due to its higher complexity, accompanied by relatively high processing demands. Interleaving task types may have further increased the processing demands, so that students' benefit from interleaving task types decreased.

With respect to the theoretical framework (see section 2.2), interleaving of task types might most likely be characterized as supporting representational sense-making processes. As argued, interleaving task types (and blocking of graphical representations) allows students to abstract across different applications of the same graphical representation to a number of different task types. The fact that the results show advantages of interleaving task types on the representational accuracy scale, which assessed students' conceptual understanding of graphical representations of fractions, including their ability to transfer this knowledge to novel graphical representations not included in the Fractions Tutor, supports this interpretation.

An alternative interpretation might be that the interleaving of task types (while blocking graphical representations) supports representational fluency-building processes as students repeatedly reactivate task-specific applications of the given graphical representation. Repeated reactivation should increase the likelihood that students can retrieve that task-specific knowledge about a given graphical representation later on. The finding that the int-types condition outperforms the int-reps condition on the representational efficiency scale, which takes into account not only the accuracy but also the speed with which students can apply their understanding of graphical representations to solve fractions tasks, supports this interpretation. Note that it is possible that interleaving task types enhances both representational sense-making processes and representational fluency-building processes, so these two interpretations may not be mutually exclusive.

One limitation with respect to these interpretations of the results in terms of the theoretical framework is that representational understanding, representational fluency, and conceptual knowledge of fractions were not assessed separately. To put this interpretation to a true test, one should empirically evaluate the following hypotheses: one would expect that the int-types condition outperforms the int-reps condition on a test of reproduction of representational understanding (or representational fluency), and that this difference in representational understanding (or representational fluency) would explain an hypothesized advantage of the int-types condition over the int-reps condition on a test of conceptual knowledge of fractions that does not involve graphical representations. Such a study could then also distinguish which learning process accounts for the advantage of interleaving task types (and blocking graphical representations): representational understanding, representational fluency, or both.

Another limitation of Experiment 2 is that it does not allow to tease apart the effects of blocking versus interleaving task types independently from the effects of interleaving versus blocking graphical representations. In other words, it is unclear whether the advantage of the int-types condition over the int-reps can be attributed to the fact that task types were interleaved (rather than blocked) or to the fact that graphical representations were blocked (rather than interleaved).

In spite of these open questions regarding the mechanisms underlying the results, Experiment 2 leads to conclusions that are of both practical and theoretical importance. The results provide guidance for developers of learning materials that include multiple graphical representations that are used across a variety of task types. Based on the findings of Experiment 2, I recommend that they interleave task types and block representations in order to promote conceptual understanding of graphical representations. Furthermore, Experiment 2 extends prior research on interleaved practice by showing that the dimension on which learning tasks are interleaved matters: interleaving task types benefits the acquisition of robust conceptual knowledge about representations more than interleaving of graphical representations does. This finding extends the literature on learning with multiple representations by demonstrating that the temporal sequence of multiple representations has an effect on students' robust learning. Whether the interpretation that interleaving learning tasks along the dimension with greatest variability holds as a general principle, remains to be confirmed by future research.

4.3 Experiment 3: Interleaving graphical representations

Experiment 2 contrasts the effects of interleaving task types (while blocking graphical representations) and interleaving graphical representations (while blocking task types). The results from Experiment 2 show that interleaving task types has a larger effect on students' robust learning of conceptual knowledge of fraction representations than interleaving graphical representations. However, it remains an open question whether interleaving graphical representations (compared to blocking graphical representations) while using a constant practice schedule across conditions can further enhance students' robust learning of fractions. Experiment 3 investigates this question by contrasting different practice schedules of graphical representations, while constantly using a moderately interleaved schedule for task types.

In this section, I first describe results with respect to students' learning outcomes. I then describe additional findings obtained from a small-scale think-aloud study and from the analysis of tutor log data. I discuss the practical and theoretical implications from this experiment with regard to the practice schedules and multiple representations literature as well as a possible interpretation of the results in the light of my theoretical framework for learning with multiple graphical representations.

4.3.1 Classroom experiment: effects of practice schedules on learning outcomes

When designing instruction that uses multiple graphical representations, curriculum designers must decide how to temporally sequence the different graphical representations. How frequently should the curriculum alternate between graphical representations? Practice schedules are likely to have an impact on students' understanding of individual representations and on their understanding of connections between different representations and, consequently, how well they learn the underlying mathematical concepts. In particular, it may matter whether the different representations are practiced in a "blocked" manner (e.g., circle – circle – circle – number line – number line – number line) or are interleaved with practice of other representations (e.g., circle – number line – circle – number line – circle – number line). As argued in Experiment 2, interleaved practice schedules provide frequent opportunities to reactivate knowledge that differs between subsequent learning tasks, and to abstract across these learning tasks. Interleaved practice with graphical reprsentations may allow students to reactivate knowledge specific to the given graphical representation, thereby strengthening that knowledge. Furthermore, by providing frequent opportunities for students to compare different graphical representations to one another (every time the student switches from one representation to the other), interleaving graphical representations might allow students to make connections between different representations. Research in a variety of domains shows that interleaved practice schedules lead to better long-term retention and transfer than blocked practice (Battig, 1972; de Croock et al., 1998; Helsdingen et al., 2011; Pashler et al., 2007; Rohrer & Taylor, 2007; Schmidt & Bjork, 1992; Schneider, 1985; Simon & Bjork, 2001; van Merriënboer et al., 2002). A limitation of this research is that it has exclusively focused on practice schedules of different task types. Whether or not the finding that interleaved practice schedules lead to better learning generalizes to practice schedules of multiple graphical representations is an open question. Yet, as mentioned, whether to present multiple graphical representations in a blocked or interleaved manner is an important design decision that developers face in virtually any STEM domain.

4.3.1.1. Research questions and hypotheses

Experiment 3 addresses the question: does interleaving graphical representations (in addition to moderately interleaving task types) enhance students' robust learning of fractions?

Specifically, Experiment 3 investigates the following hypotheses:

Hypothesis 1: Interleaved practice schedules of multiple graphical representations (while also moderately interleaving task types) enhance students' learning of robust knowledge of fractions.

Hypothesis 2: Multiple graphical representations, when provided in an interleaved fashion (while also moderately interleaving task types), lead to better learning than a single graphical representation.

4.3.1.2. Methods

To investigate these hypotheses, I conducted a classroom experiment that contrasted the effects of different practice schedules of multiple graphical representations on students' robust learning of fractions, while moderately interleaving task types (constantly across all conditions).

Experimental design

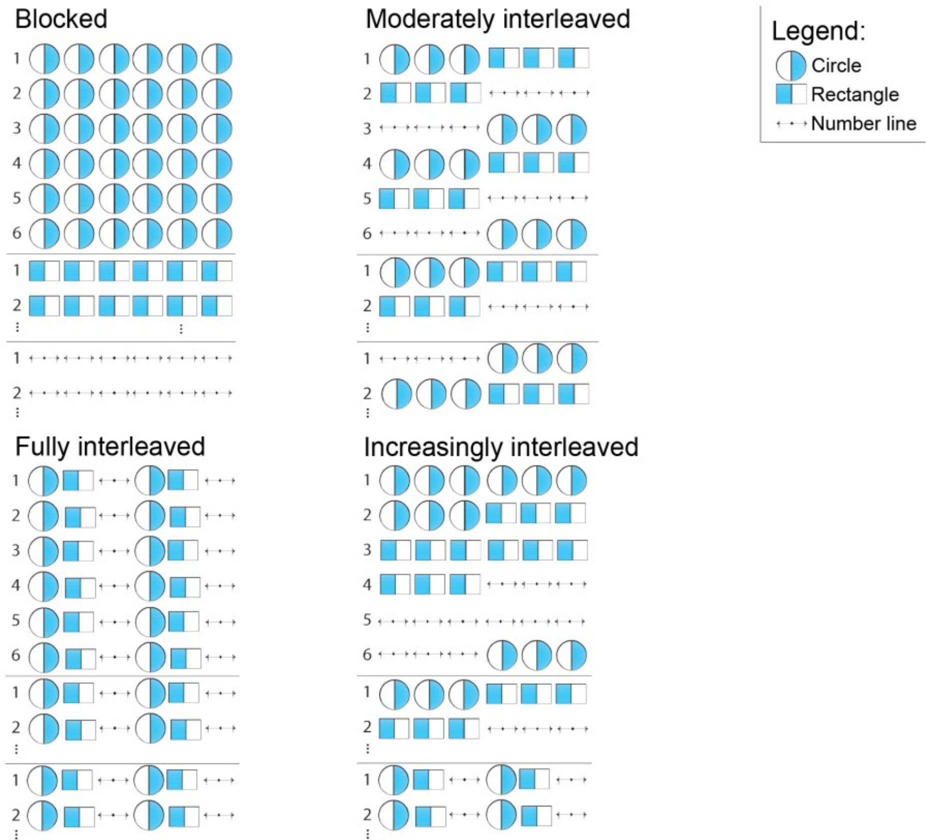


Fig. 22. Experiment 3: Practice schedules for the multiple graphical representations conditions. In all conditions, six task types were presented three times. Numbers 1-6 indicate task types, shapes depict representations.

In all conditions, students worked through the same sequence of task types and fraction problems, and switched task types after every six of a total of 108 problems. Each task type was visited three times. A moderately interleaved schedule for task types was chosen so as to prevent the possible scenario that subsequent learning tasks are too dissimilar, which (as argued in Experiment 2), might hamper students' ability to abstract common knowledge across learning tasks. I randomly assigned students to one of five conditions. In the *blocked* condition, students switched graphical representations after 36 problems. In the *moderate* condition, students switched representations after every six problems. In the *fully interleaved* condition, students switched representations after each problem. In the *increased* condition, the length of the blocks was gradually reduced from twelve problems at the beginning to a single problem at the end. To account for possible effects of the order of graphical representations, I randomized several plausible orders in which students encountered the graphical representations (i.e., students never received the number line first, since the number line is the hardest graphical representation, possible orders of representations were circle – rectangle – number line, circle – number line – rectangle, rectangle – circle – number line, rectangle – number line – circle). Finally, students in the single graphical representation conditions were randomly assigned to work on all tutor problems with only the circle, the rectangle, or the number line, respectively.

Participants

Experiment 3 was conducted with 587 4th- and 5th-grade students from six schools (31 classes). I excluded students who missed at least one test day, and who completed less than 67% of all tutor problems (students in the blocked condition switched representations after problem 36, and 72 out of 108). I had to apply this stringent criterion to ensure that students in the blocked condition encountered all three graphical representations (see Fig. 22). This results in a total of N = 290 (n = 63 in blocked, n = 53 in moderately interleaved, n = 52 in fully interleaved, n = 62 in increasingly interleaved, n = 60 in the single-representation conditions [n = 21 in single-circle, n = 20 in single-rectangle, n = 19 in single-number-line]). Table 10 shows the number of included and excluded students per condition.

	included	**excluded**
blocked	63	58
moderately interleaved	53	60
fully interleaved	52	64
increasingly interleaved	62	62
single-circle	21	17
single-rectangle	20	16
single-number-line	19	20

Table 10. Experiment 3: Number of students included and excluded by condition.

Procedure

Experiment 3 took place at the end of the school year 2009/20010. Students' regular math teachers led the sessions, but researchers were present in the classrooms at all times to assist teachers in answering questions specific to the use of the Fractions Tutor.

Prior to working with the Fractions Tutor, students completed a pretest. The pretest took about 30 minutes. On the following day, all students started working with the Fractions Tutor (with the version of the Fractions Tutor depicted in Fig. 10, see section 3.3.3.3, on topics 1-5 in Table 1, see section 3.4). Students accessed the tutoring system from the computer lab at their schools and worked on the Fractions Tutor for about five hours as part of their regular math instruction for five to six consecutive school days (depending on the length of the respective school's class periods). All students worked on the Fractions Tutor at their own pace, but the time students spent with the system was held constant across classrooms and across experimental conditions. On the day following the tutoring sessions, students completed the immediate posttest, which took about 30 minutes. Seven days after the posttest, students completed an equivalent delayed posttest.

Test instruments

I assessed students' knowledge of fractions at three test times using three equivalent test forms. I randomized the order in which they were administered. The tests included four knowledge types: reproduction with area models (i.e., circles and rectangles), reproduction with number lines, conceptual transfer and procedural transfer. The area model items and number line items covered identifying fractions given a graphical representation, making a graphical representation given a symbolic fraction, and recreating the unit given a graphical representation of both unit fractions and proper fractions, as covered by the Fractions Tutor. Conceptual transfer items included proportional reasoning questions with and without graphical representations. Procedural transfer items included comparison questions with and without graphical representations. The changes made to the tests with respect to the tests used in Experiment 2 reflect the changes made to the Fractions Tutor, since each experiment was part of the iterative tutor development process described in section 3.3. Specifically, the reproduction with area models and reproduction with number line scales correspond to the reproduction items in the representational knowledge and operational knowledge scales of the test used in Experiment 2, while aligning with the revised version of the Fractions Tutor. The conceptual transfer scale of the test corresponds to the transfer items that were included in the representational knowledge scale of the test used in Experiment 2, again reflecting changes made to the Fractions Tutor between the two experiments. The procedural transfer scale corresponds to the transfer items that were part of the operational knowledge scale of the test used in Experiment 2, again while being adapted to reflect changes made to the Fractions Tutor. The theoretical structure of the test (i.e., the four knowledge types just mentioned) resulted from a factor analysis performed on the pretest data. Test items including the number line seemed to be more challenging for students than area models. Examples of the test items for each of the four knowledge types can be found in Appendix 6.

4.3.1.3. Results

As mentioned, I analyzed the data of $N = 290$ students. There was no significant difference between conditions with respect to the number of students excluded ($\chi^2 < 1$). There were no significant differences between conditions at pretest for any dependent measure, *p*s > .10. There was no significant effect for order of multiple graphical representations for any dependent measure, $F(5, 285) = 1.56$, *p*s > .10.

I used an HLM (see Raudenbush & Bryk, 2002) with four nested levels to analyze the data. At level 1, I modeled performance on each of the tests for each student. At level 2, I accounted for differences between students. Level 3 models random differences between classes, and level 4 random differences between schools. The HLM is the outcome of a forwards-inclusion procedure in which I used the BIC to find whether the inclusion of a variable increased model fit. If the BIC decreased as a consequence of including a variable (indicating better model fit), I kept the variable. If the BIC did not decrease, I did not include the variable. I tested a number of variables, including teacher, sequence of graphical representations, test form sequence, grade level, number of problems completed, total time spent with the tutor, random intercepts and slows for classes and schools. Equation 3 shows the resulting HLM:

$$Y_{ijkl} = (((\mu + W_l) + V_{kl}) + \beta_3 * c_j + \beta_4 * p_j + \beta_5 * c_j * p_j + U_{jkl}) + \beta_1 * t_i + \beta_2 * c_j * t_i + R_{ijkl} \quad (3)$$

with

(level 1) $Y_{ijkl} = \varepsilon_{jkl} + \beta_1 * t_i + \beta_2 * c_j * t_i + R_{ijkl}$

(level 2) $\varepsilon_{jkl} = \delta_{kl} + \beta_3 * c_j + \beta_4 * p_j + \beta_5 * c_j * p_j + U_{jkl}$

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

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

with the index *i* standing 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 effect of the interaction of condition c_j with test time t_i, β_3 is the parameter for the effect of condition c_j, β_4 is the parameter for the effect of student$_j$'s performance on the pretest p_j, β_5 is the parameter for an aptitude-treatment interaction between condition c_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.

Since the HLM described in (3) uses students' pretest scores as a covariate, it does not allow us to analyze whether students in the various conditions improved from pretest to immediate and delayed posttest. To analyze learning gains, I excluded pretest score in the dependent variable, as well as an interaction of pretest with condition, yielding:

$$Y_{ijkl} = (((\mu + W_l) + V_{kl}) + \beta_3 * c_j + U_{jkl}) + \beta_1 * t_i + \beta_2 * c_j * t_i + R_{ijkl} \quad (4)$$

with

(level 1) $Y_{ijkl} = \varepsilon_{jkl} + \beta_1 * t_i + \beta_2 * c_j * t_i + R_{ijkl}$

(level 2) $\varepsilon_{jkl} = \delta_{kl} + \beta_3 * c_j + U_{jkl}$

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

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

with the index *i* standing for test time (i.e., pretest, immediate, and delayed posttest). The dependent variable Y_{ijkl} is student$_j$'s score on the dependent measures at test time t_i (i.e., pretest, immediate posttest, or delayed posttest). Excluded were the parameters β_4 for the effect of student$_j$'s performance on the pretest p_j, and the parameter β_5 is for an aptitude-treatment interaction between condition c_j and student$_j$'s performance on pretest p_j.

I used planned contrasts and posthoc comparisons, all of which were computed as part of the HLM to clarify results from the HLM analysis. All reported *p*-values were adjusted using the Bonferroni correction for multiple comparisons.

Effects of practice schedules on students' learning

Table 11 shows the means and standard deviations for the dependent measures by condition and test time.

To investigate hypothesis 1 (that interleaved practice schedules enhance students' learning of robust knowledge of fractions), I computed the HLM presented in formula (3) for only the multiple graphical representations conditions. There was no significant main effect of condition on any knowledge type, indicating that there was no global effect of practice schedules of multiple graphical representations across immediate and delayed posttests. An interaction between test time and condition was marginally significant for reproduction with area models, $F(3, 867) =$

2.57, $p < .10$, $\eta^2 = .01$, so that the effect of condition was stronger on the immediate posttest than on the delayed posttest, indicating that the effect of practice schedules depends on test time. The interaction between pretest score and condition was marginally significant for conceptual transfer, $F(3, 219) = 2.52$, $p < .10$, $\eta^2 = .02$, demonstrating that students with different pretest scores benefit from different practice schedules.

		Reproduction with area models	Reproduction with number lines	Conceptual transfer	Procedural transfer
Pretest	Blocked	.56 (.26)	.45 (.28)	.60 (.30)	.51 (.34)
	Moderately interleaved	.64 (.25)	.50 (.27)	.70 (.27)	.51 (.31)
	Fully interleaved	.59 (.31)	.46 (.25)	.71 (.22)	.60 (.36)
	Increasingly interleaved	.64 (.25)	.48 (.26)	.69 (.25)	.55 (.36)
	Single-circle	.58 (.37)	.47 (.29)	.59 (.28)	.49 (.31)
	Single-rectangle	.55 (.31)	.52 (.26)	.67 (.28)	.50 (.37)
	Single-number-line	.65 (.26)	.44 (.26)	.69 (.25)	.41 (.32)
Immediate posttest	Blocked	.62 (.25)	.51 (.31)	.72 (.30)	.52 (.37)
	Moderately interleaved	.64 (.26)	.54 (.25)	.77 (.24)	.55 (.36)
	Fully interleaved	.71 (.24)	.58 (.25)	.78 (.23)	.60 (.34)
	Increasingly interleaved	.73 (.22)	.58 (.26)	.75 (.27)	.53 (.37)
	Single-circle	.61 (.28)	.50 (.32)	.70 (.30)	.55 (.36)
	Single-rectangle	.67 (.25)	.53 (.32)	.77 (.27)	.48 (.40)
	Single-number-line	.67 (.27)	.57 (.28)	.70 (.32)	.40 (.36)
Delayed posttest	Blocked	.69 (.25)	.57 (.30)	.71 (.31)	.60 (.37)
	Moderately interleaved	.71 (.20)	.62 (.25)	.78 (.26)	.65 (.32)
	Fully interleaved	.69 (.25)	.64 (.25)	.83 (.19)	.56 (.37)
	Increasingly interleaved	.77 (.20)	.60 (.28)	.77 (.25)	.55 (.32)
	Single-circle	.72 (.29)	.57 (.31)	.74 (.26)	.63 (.36)
	Single-rectangle	.65 (.27)	.52 (.32)	.73 (.36)	.54 (.33)
	Single-number-line	.61 (.30)	.50 (.30)	.75 (.29)	.40 (.30)

Table 11. Experiment 3: Means and standard deviations (in parentheses) for dependent measures at pretest, immediate posttest, delayed posttest by condition.

To clarify the interaction between test time and condition, I used posthoc contrasts separately for the immediate and the delayed posttest. To limit the number of comparisons, I only compared the most successful multiple graphical representations condition (determined separately for each measure) against the remaining three multiple graphical representations conditions taken together, as summarized in Table 12: for reproduction with area models, the increasingly interleaved condition was the most successful one. For both reproduction with number lines and conceptual transfer, the fully interleaved condition was the best, and for procedural transfer, the moderately interleaved condition was the best. I found some support for a benefit of interleaving multiple graphical representations: the fully interleaved condition significantly outperformed the not-fully-interleaved conditions (i.e., blocked, moderately interleaved, and increasingly interleaved) on conceptual transfer at the delayed posttest. Furthermore, I found a marginally significant advantage for the increasingly interleaved condition over the not-increasingly-interleaved conditions (i.e., blocked, moderately interleaved, and fully interleaved) on reproduction with area models at the immediate and the delayed posttests.

Effect	Test	Reproduction with area models	Reproduction with number lines	Conceptual transfer	Procedural transfer
fully interleaved > blocked, moderately interleaved, increasingly interleaved	post	-	ns	ns	-
	delayed	-	ns	$p < .05$, $d = .33$	-
increasingly interleaved > blocked, moderately interleaved, fully interleaved	post	$p < .10$, $d = .30$	-	-	-
	delayed	$p < .10$, $d = .30$	-	-	-
moderately interleaved > blocked, fully interleaved, increasingly interleaved	post	-	-	-	ns
	delayed	-	-	-	ns

Table 12. Experiment 3: Results from posthoc comparisons on differences between multiple representations conditions at immediate posttest (post) and delayed posttest (delayed) by type of knowledge. “ns” indicates non-significant differences. “-” indicates that no posthoc comparisons were computed.

To clarify the interaction between pretest score and condition on conceptual transfer, I computed posthoc comparisons for students with extremely low or high pretest scores. For students with a low pretest score of 15%, 20%, and 25%, I found a significant advantage for the fully interleaved over the blocked condition (*p*s < .05). I found no differences between conditions for the high prior knowledge students.

Effects of practice schedules on students’ benefit from multiple graphical representations

To investigate hypothesis 2 (that multiple graphical representations, when provided in an interleaved fashion, lead to better learning than a single graphical representation), I applied the HLM described in formula (3) for the fully interleaved condition and the single-representation control conditions. The fully interleaved condition was selected for this analysis because it was the most successful condition for two of four measures (see Table 11), especially for students with low prior knowledge. I then computed planned contrasts that compared the fully interleaved condition to the single graphical representation conditions for each knowledge type at the immediate posttest and at the delayed posttest. I found a significant advantage for the interleaved condition over the single-representation conditions on reproduction with number lines at the immediate posttest, $t(445) = 2.09$, $p < .05$, $d = .09$, on reproduction with number lines at the delayed posttest, $t(445) = 2.66$, $p < .01$, $d = .12$, and on transfer of conceptual knowledge at the delayed posttest, $t(445) = 2.27$, $p < .05$, $d = .10$.

Condition	Effect	Reproduction with area models	Reproduction with number lines	Conceptual transfer	Procedural transfer
blocked	post > pre	ns	ns	$p < .05$, $d = .42$	ns
	delayed > pre	$p < .05$, $d = .52$	$p < .01$, $d = .39$	$p < .05$, $d = .39$	ns
moderately interleaved	post > pre	ns	ns	$p < .05$, $d = .29$	ns
	delayed > pre	ns	$p < .01$, $d = .50$	$p < .05$, $d = .30$	$p < .05$, $d = .45$
fully interleaved	post > pre	$p < .05$, $d = .45$	$p < .01$, $d = .51$	$p < .01$, $d = .34$	ns
	delayed > pre	$p < .05$, $d = .38$	$p < .01$, $d = .75$	$p < .01$, $d = .60$	ns
increasingly interleaved	post > pre	$p < .05$, $d = .38$	$p < .01$, $d = .43$	ns	ns
	delayed > pre	$p < .05$, $d = .55$	$p < .01$, $d = .46$	ns	ns
single-circle	post > pre	ns	ns	ns	ns
	delayed > pre	ns	ns	ns	ns
single-rectangle	post > pre	ns	ns	ns	ns
	delayed > pre	ns	ns	ns	ns
single-number-line	post > pre	ns	ns	ns	ns
	delayed > pre	ns	ns	ns	ns

Table 13. Experiment 3: Results on test scores at immediate posttest (post) over pretest (pre) and delayed posttest (delayed) over pretest by knowledge types and conditions. “ns” indicates non-significant differences.

To further investigate whether students’ learning gains differ between conditions, I analyzed learning gains using the HLM described in formula (4). The main effect of test time was significant for reproduction with number lines, $F(2, 867) = 20.09$, $p < .01$, partial $\eta^2 = .03$, for reproduction with area models, $F(2, 867) = 17.54$, $p < .01$, $\eta^2 = .02$, conceptual transfer, $F(2, 867) = 38.78$, $p < .01$, partial $\eta^2 = .03$, and marginally significant for procedural transfer, $F(2, 867) = 2.84$, $p < .10$, partial $\eta^2 = .01$. The interaction between test time and condition was significant for reproduction with area models $F(12, 862) = 2.06$, $p < .05$, partial $\eta^2 = .01$. These results show that students (regardless of condition) benefited from working with the Fractions Tutor on reproduction with number lines, reproduction with area models, procedural and conceptual transfer, albeit with small effect sizes. On reproduction with area models, students’ learning gains depended on the condition.

To further clarify these results, I computed posthoc comparisons that contrasted students’ scores at the immediate posttest and the delayed posttest, compared to the pretest. Table 13 provides a summary of these posthoc comparisons. Generally, I found significant learning gains at the delayed posttest for most of the multiple graphical representations conditions on reproduction with area models, reproduction with number lines, and conceptual transfer. On procedural transfer, only the moderate condition showed significant learning gains at the delayed posttest. Finally, I found no significant learning gains for the single graphical representation conditions.

Please also refer to Rau, Rummel, and colleagues (2012) for the results from Experiment 3.

4.3.1.4. Summary

The results provide limited support for hypothesis 1, that interleaving graphical representations (while also moderately interleaving task types) enhances students' learning of robust knowledge of fractions. The blocked condition never outperformed any of the interleaved conditions. Furthermore, there was an 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. There was also a marginally significant advantage for the increasingly interleaved condition over the blocked, moderately interleaved, and fully interleaved conditions on reproduction with the number line at the immediate and the delayed posttests. Although a comparison of all multiple graphical representations conditions and the single graphical representation condition was not significant for the majority of dependent measures, the fully interleaved condition significantly outperformed the single-representation control condition on reproduction with the number line, conceptual transfer, and marginally on procedural transfer. The lack of differences on transfer of procedural knowledge may reflect the fact that the Fractions Tutor focuses on conceptual learning of fractions more so than on procedural learning.

Furthermore, the results provide some support for hypothesis 2, that students who work with interleaved practice schedules of multiple graphical representations outperform students who work with only a single graphical representation. Both on reproduction with number lines and on transfer of conceptual knowledge, the fully interleaved condition significantly outperformed the single-representation control group.

The analysis of students' learning gains provides additional support for hypothesis 2. There were significant overall learning gains only for students who worked with multiple graphical representations. Students performed significantly better on all knowledge types, although the learning gains on transfer of procedural knowledge were only marginally significant. These gains persisted until one week after the study when students completed the delayed posttest. The fact that students' performance on procedural transfer did not improve to the same degree as on the other knowledge types may again reflect the Fractions Tutor's focus on conceptual learning over procedural learning. At the same time, students who worked with only a single graphical representation did not demonstrate significant learning gains. This finding demonstrates the importance of providing students with a variety of graphical representations, each of which emphasize a particular conceptual view on fractions, to promote their learning. Experiment 3 replicates the finding from Experiment 1: multiple graphical representations can enhance students' robust learning of fractions. While the single-representation condition in Experiment 1 used only number lines, Experiment 3 extends this finding by showing that the advantage of multiple graphical representations holds also when compared to learning with only a circle, or only a rectangle representation.

In conclusion, the results from Experiment 3 provide some support for the notion that multiple graphical representations should be provided in an interleaved fashion (in addition to moderately interleaving task types). These results extend the findings from Experiment 2, that interleaving task types (while blocking graphical representations) enhances students' acquisition of robust conceptual knowledge of fraction representations. Although Experiments 2 and 3 are not directly comparable, Experiment 3 suggests that the advantage of the int-types condition over the int-reps condition in Experiment 2 was not due to blocking graphical representations, but due to interleaving task types. Taken together, Experiment 3 extends prior research on practice schedules of task types by showing that practice schedules of graphical representations have an impact on students' learning.

In conclusion, the findings from Experiment 3 provide (albeit limited) support for the notion that instructional materials should provide interleaved practice with multiple graphical representations, if the goal is to promote the acquisition of robust conceptual knowledge that can transfer to novel tasks.

4.3.2 Think-aloud study to investigate learning processes

What might be the mechanisms by which interleaving graphical representations lead to more robust learning of domain knowledge? As discussed in Experiment 2, one possible mechanism is *repeated reactivation of learning content*. When working with interleaved task types, students have to reactivate the knowledge needed to solve each learning task more often than when working with blocked learning tasks (de Croock et al., 1998; Lee & Magill, 1983, Sweller, 1990). It is assumed that students store knowledge about how to solve a given task type in long-term memory. When working on a task type that requires this knowledge, students load it into working memory. This process of retrieving knowledge from long-term memory and loading it into working memory strengthens the association between the task type and the knowledge, which increases the likelihood that students can recall that knowledge later on (e.g., Anderson, 1993, 2002).

Another explanation for the advantage of interleaved practice schedules over blocked practice schedules is that they help students to *abstract knowledge* from different task types presented consecutively (de Croock et al., 1998; Shea & Morgan, 1979). Abstraction occurs when students make connections between the knowledge required by different task; for instance, by comparing properties of the different tasks to one another. This process requires that knowledge from the previous task is still active in working memory when the knowledge required for the subsequent task is loaded into working memory. As students make connections between the knowledge required to solve different tasks, students may see more clearly what properties are incidental and which are key to solving a given task. This connection-making process allows students to process the features relevant to knowledge construction and allows them to abstract the knowledge common to consecutive tasks (Paas & van Gog, 2006; van Merriënboer et al., 2002). Students more frequently encounter dissimilar tasks back to back in interleaved practice schedules, but less frequently in blocked practice schedules.

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 multiple graphical representations across consecutive tutor problems, and whether students' ability to make these connections can be enhanced by prompting them to do so.

Six 5th-grade students participated in the think-aloud study. The think-aloud study was conducted in the laboratory and included three sessions. During the first session, students took the same pretest that was used in the experimental study reported above. The pretest took about 30 minutes to complete. During the second session, students worked for one hour on a subset of problems taken from the interleaved version of the tutoring system while being prompted to think aloud, following the procedure described in Ericsson and Simon (1984). In the third session, students worked with similar tutor problems for one hour while being prompted to relate the different graphical representations to one another. I varied the type of prompts based on a within-subjects design: the prompt questions were either implicit (i.e., without directly prompting comparisons between the representations; e.g. "How is this problem the same as the last two you did?" or "How is this problem different from the last one you did?"), or explicit (i.e., directly referring to aspects that the different representations share; e.g., "What is the unit in the circle / rectangle / number line?" or "How are the rectangle and the circle and the number line the same / different?"). All students received two implicit prompts and four explicit prompts, in a fixed sequence.

Students' utterances were recorded and transcribed. I combined top-down and bottom-up approaches in developing a coding scheme: I identified types of connections that students might make prior to the think-aloud study, and then refined the coding scheme after viewing the transcripts from the think-aloud study. Connections between graphical representations were coded as surface connections if they either referred to the color of the representation, the shape of the representation, or the action performed on the representation (e.g., dragging and dropping). For example, when asked "how is the circle like the rectangle?", a student's response "you have to drag something into a diagram of the unit" would be coded as a surface connection. Connections were coded as conceptual if they referred to the corresponding features of the representations (i.e., numerator, denominator, unit), or the magnitude represented. For instance, when asked: "how is the number line like the circle?" for improper fractions, a student's answer "they both have one whole unit plus a fraction of another unit that's the same" would be coded as a conceptual connection.

The results from the pretest indicate that all students had a good understanding of fractions. During the spontaneous comparison phase of the think-aloud study, I found only five instances of connections. These five connections were uttered by five of the six students. All five connections were surface connections. In addition to these spontaneous connections, I found 138 instances of prompted connection making. Table 14 summarizes the average number of connections coded as surface and conceptual connections for implicit and explicit prompts. Given the small number of students, a statistical test on the types of connections in response to implicit and explicit prompts is not warranted. The data in Table 14 suggest, however, that students generated substantially more surface connections than conceptual connections. We also see that the implicit prompts yielded most of the surface connections, but few of the conceptual connections. Explicit prompts seem to have yielded more of the conceptual connections, compared to the implicit prompts.

	Implicit prompts	**Explicit prompts**	***Overall***
Surface	4.17	2.33	*2.94*
Conceptual	.58	1.63	*1.28*
Overall	*2.38*	*1.98*	

Table 14. Experiment 3: Number of surface connections and conceptual connections by implicit and explicit prompts averaged across students.

In sum, results from the think-aloud study show that students tend not to spontaneously make connections between multiple graphical representations: I found only five spontaneous connections, and all of them were surface connections. However, students are able to make these connections when prompted to do so. In particular, explicit prompts are well-suited to enhance conceptual connections.

With regards to learning mechanisms, these findings suggest that the advantage of interleaved practice reported above does not stem from spontaneous connection-making activities between multiple graphical representations. Thus, the benefit from interleaved practice with multiple graphical representations does not seem to stem from conscious abstraction across the different representations. Rather, interleaved practice may be attributed to requiring students to repeatedly reactivate knowledge about the specific graphical representations. The fact that students were able to make connections when prompted to do so demonstrates that the lack of spontaneous connection-making activities is not an artifact of the think-aloud method being an unsuitable metric for detecting students' connection-making processes.

Please also refer to Rau, Rummel, and colleagues (2012) for the results from the think-aloud study.

4.3.3 Educational data mining to investigate learning processes

Task type	**Blocked**	**Moderate**	**Interleaved**	**Increased**
1	.85 (.08)	.86 (.06)	.86 (.06)	.87 (.05)
2	.88 (.07)	.89 (.07)	.89 (.05)	.89 (.05)
3	.91 (.05)	.88 (.07)	.88 (.06)	.87 (.07)
4	.87 (.06)	.84 (.07)	.81 (.07)	.84 (.06)
5	.83 (.10)	.86 (.06)	.83 (.08)	.83 (.07)
6	.88 (.10)	.90 (.06)	.89 (.07)	.89 (.07)
Overall	.87 (.08)	.87 (.07)	.86 (.07)	.87 (.06)

Table 15. Experiment 3: Results on average number of correct first attempts by task type and condition (standard deviation in brackets). Higher numbers indicate higher performance during the acquisition phase.

To further investigate the mechanisms underlying the advantage of the fully interleaved condition over the blocked condition, I augment the findings from the traditional analysis of pretest and posttest data by applying a knowledge tracing algorithm (Corbett & Anderson, 1995) to the log data obtained from the Fractions Tutor. Analyzing student performance during the acquisition phase (i.e., while students learn) is particularly interesting when investigating the effects of practice schedules: a common finding is that interleaved practice schedules lead to better long-term retention and to better transfer than blocked schedules, but they often lead to worse performance during the acquisition phase (Battig, 1972; de Croock et al., 1998; Helsdingen et al., 2011; Pashler et al., 2007; Rohrer & Taylor, 2007; Schmidt & Bjork, 1992; Schneider, 1985; Simon & Bjork, 2001; van Merriënboer et al., 2002). Therefore, it is often believed that the advantage of interleaved practice over blocked practice is not apparent during the acquisition phase, but can only be detected with long-term retention tests and transfer tests administered *after* the acquisition phase.

Table 15 provides a summary of students' performance on the Fractions Tutor problems during the acquisition phase, based on the overall first-attempt correct steps students made during practice with the Fractions Tutor. A repeated measures ANOVA with students' performance on each task type as the dependent measure and practice schedule as the independent factor showed that students' performance during the acquisition phase did not significantly differ between practice schedules ($F < 1$). Planned contrasts between the blocked condition and each of the interleaved condition did not yield significant differences in students' performance (ts < 1). Therefore, based on raw performance measures, I do not find evidence that interleaved practice with graphical representations promotes students' learning during the acquisition phase. In the light of the literature on practice schedules, this lack of an advantage of the interleaved condition on correct attempts is not surprising.

To investigate whether latent measures of learning can detect an advantage of interleaved over blocked practice during the acquisition phase, I used a Bayesian Network model based on knowledge tracing (Corbett & Anderson, 1995). Knowledge tracing uses a two state Hidden Markov Model assumption of learning, which uses correct and incorrect responses in students' problem-solving attempts to infer the probability of a student knowing the skill underlying the problem-solving step at hand. I combined this model with several other extensions to knowledge tracing to each of the four experimental conditions of the experimental study to investigate differences in model learning rates between the conditions in the Fractions Tutor.

4.3.3.1. Bayesian Knowledge Tracing models

To this end, I evaluated four Bayesian Knowledge Tracing models based on the Fractions Tutor log data. Two of the models were created for the purpose of analyzing the learning rates of the conditions in the experiment while the other two were used as baseline models to gauge the relative predictive performance of the new models. None of the tested models included a knowledge component model, so each step in the tutor is treated as a knowledge component.

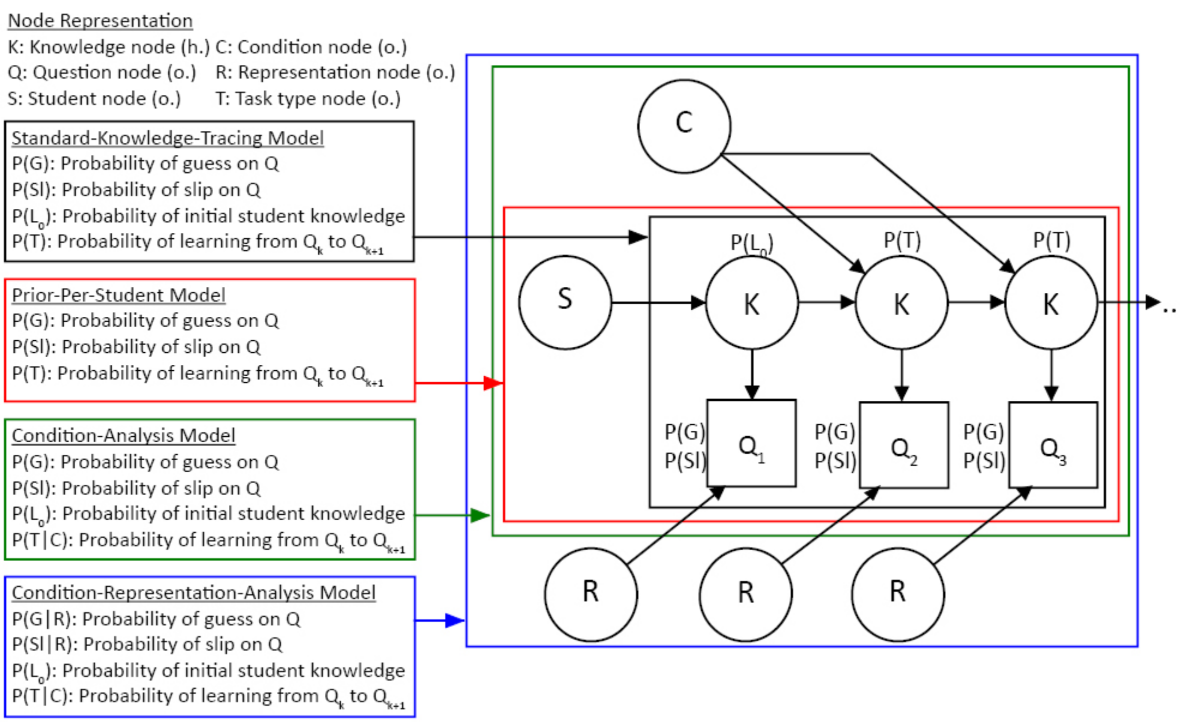


Fig. 23. Experiment 3: Overview of the four different Bayesian Networks tested, with observed (o.) and hidden (h.) nodes.

I employed two models that served as benchmarks for model fit and designed two novel models for evaluating learning differences among the experiment conditions. I compared the resulting four Bayesian models all of which were based around knowledge tracing. Fig. 23 provides an overview of the different models that were compared. The Standard-Knowledge-Tracing model and the Prior-Per-Student model correspond to the two benchmark models. The Standard-Knowledge-Tracing model includes only knowledge tracing without taking students' prior knowledge (S) (Pardos & Heffernan, 2010), experimental condition (C), or fraction representation (R) into account. The Prior-Per-Student model (Pardos & Heffernan, 2010) includes the students' individualized prior knowledge (S). Both the Standard-Knowledge-Tracking model and the Prior-Per-Student model assume that there is a probability that a student will transition from the unlearned to the learned knowledge state at each opportunity regardless of the particular problem just encountered or practice schedule of the student.

The Condition-Analysis model and the Condition-Representation-Analysis models serve as a means to answer hypothesis 1 as stated for the analysis of learning outcome, that interleaved practice schedules of multiple graphical representations enhance students' learning. Hence, I depart from the simplifying assumption of a single learning rate per skill and instead fit a separate learning rate for each of the four practice schedules implemented in the Fractions Tutor. To do so, I adapted modeling techniques from prior work, which evaluated the learning value of different forms of tutoring in (non-experiment) log data of an intelligent tutor (Pardos et al., 2010). Specifically, I estimated four different learning rates per task type, each corresponding to the particular condition (i.e., blocked practice, fully interleaved, moderately interleaved, or increasingly interleaved) assigned to the student – as opposed to using a single learning rate per task type. The Condition-Analysis model includes students' prior knowledge and models the effect of experimental condition (C). Finally, the Condition-Representation-Analysis model incorporates students' prior knowledge (S), condition (C), and the graphical representation encountered by each student in each problem (R). Specifically, I hypothesized that the different learning rate estimates will significantly differ between experimental conditions, within each given task type (in the Condition-Analysis model), and between graphical representations (in the Condition-Representation-Analysis model).

To model different learning rates within knowledge tracing, I adapted modeling techniques from prior work, which evaluated the learning value of different forms of tutoring in (non-experiment) log data of an intelligent tutoring system (Pardos et al., 2010). Different representations of fractions are expected to result in different degrees of difficulty in solving the tutor problem (Charalambous & Pitta-Pantazi, 2007). The Condition-Representation-Analysis model used techniques from Knowledge-Tracing-Item-Difficulty-Effect Model (Pardos & Heffernan, 2011) to model different guess and slips for problems depending on the representation used in the tutor problem.

Since there was no knowledge component model, I determined model fit by task type. To evaluate predictive performance, reported below, I used a 5-fold cross-validation at the student level. For the reporting of learning rates by practice schedule, all data was used to train the model.

The parameters in all four models were fit using the Expectation Maximization algorithm implemented in Kevin Murphy's Bayes Net Toolbox (Murphy, 2001). For the Condition-Representation-Analysis Model the number of parameters fit per task was 12 (2 prior + 4 learn rate + 3 guess + 3 slip). Probabilities of knowledge were set to 1 if the skill was already known, $P(L_{n-1}) = 1$, to represent a zero chance of forgetting, an assumption made in standard knowledge tracing. If a student was previously (at learning opportunity $n - 1$) in the unlearned state, the probability that he/she will now (at opportunity n) have transitioned to the learned state between problems is:

$$P(L_{n-1}) + \left((1 - P(L_{n-1})\right) * P(T|C_s)), \tag{5}$$

where $P(L_{n-1})$ is the probability of a student already knowing the skill, C is the condition assigned to a student (i.e., blocked, fully interleaved, moderately interleaved, increasingly interleaved), and T is the given task type.

4.3.3.2. Evaluation results

Model	RMSE	AUC
Condition-Representation-Analysis Model	.3427	.6528
Standard-Knowledge-Tracing Model	.3445	.6181
Condition-Analysis Model	.3466	.5509
Prior-Per-Student Model	.3469	.5604

Table 16. Experiment 3: Summary of the cross-validated prediction results of the four tested models using RMSE and AUC metrics.

Task type	Blocked	Moderate	Interleaved	Increased
1	.0061	.0061	.0080	.0072
2	.0019	.0032	.0065	.0036
3	.0149	.0059	.0337	.0030
4	.0037	.0022	.0035	.0014
5	.0108	.0220	.0124	.0130
6	.0043	.0107	.0078	.0090
Overall	*.0062*	*.0056*	*.0120*	*.0062*

Table 17. Experiment 3: Learning rates by task type and condition from the Condition-Representation Analysis Model. Higher numbers indicate higher learning rates during the acquisition phase.

To evaluate the predictive accuracy of each of the student models mentioned above, I conducted a 5-fold cross-validation at the student level. Cross-validating at the student level increases confidence that the resulting models and their assumptions about learning will generalize to new groups of students. The metric used to evaluate the models is root mean squared error (RMSE) and Area Under the Curve (AUC). Lower RMSE equals better prediction accuracy. For AUC, a score of 0.50 represents a model that is predicting no better than chance. An AUC of 1 is a perfect prediction.

As shown in Table 16, the Standard-Knowledge-Tracing model has an overall RMSE of .3445, the Prior-Per-Student model has an RMSE of .3469, the Condition-Analysis model has an RMSE of .3466, and the Condition-Representation-Analysis model has the lowest RMSE with .3427 as well as the best AUC. These results demonstrate that the Bayesian network that includes students' prior knowledge (S), experimental condition (C), and representations used for a certain problem (R) provides the best model fit.

Table 17 shows the learning rates obtained from the Condition-Representation-Analysis model for each condition for each of the task types that the Fractions Tutor covered. Overall, the learning rate estimates align with the results obtained from the learning outcome data: the interleaved condition demonstrates higher learning rates overall than the other conditions. The learning rates by task type provide more specific information on the nature of the differences between conditions in learning rates. For all but the fourth task type (naming improper fractions), the fully interleaved condition demonstrates a higher learning rate than the blocked condition. To test whether these differences are statistically significant, I employed the binomial test used by Pardos et al. (2010). The advantage of the interleaved practice schedule over the blocked practice schedule was statistically significant for task types 1, 2 and 3 (ps $< .05$) and moderately significant for task type 5 ($p < .10$). The interleaved condition achieved the highest overall learning rate, which was twice that of any other condition. Given that, to the best of my knowledge, prior research on practice schedules has not found evidence of an advantage of interleaved practice over blocked practice during the acquisition phase, and since performance, as established by the average number of errors made during the acquisition phase (see Table 15), did not differ between conditions, this finding is remarkable.

Please consider Rau and Pardos (2012) for a more detailed discussion of these results.

4.3.3.3. Summary

The findings from the Bayesian Knowledge Tracing analysis support and augment the findings from the learning outcome data in several ways. First, the finding that the Condition-

Representation-Analysis model provides a better fit to the log data than the other models (especially those that do not include a node for the effect of condition) confirms the interpretation from the analysis of the learning outcome data that practice schedules of multiple graphical representations matter. Furthermore, the finding that the representation used in a tutor problem is a useful predictor of learning is consistent with the notion that different graphical representations provide different conceptual views on fractions in a way that influences how students understand fractions (Charalambous & Pitta-Pantazi, 2007).

Second, the learning rate estimates per condition support the hypothesis that interleaved practice schedules of multiple graphical representations of fractions lead to better learning than a blocked practice schedule. This finding is interesting, especially in light of the lack of statistically significant differences in students' performance on the Fractions Tutor during the acquisition phase. As shown in Table 15, students' performance on the Fractions Tutor problems, measured by the success rate on the first attempt on each step, does not differ between conditions. The literature on contextual interference shows that interleaved practice schedules often impair performance during the acquisition phase (de Croock et al., 1998; Immink & Wright, 1998; Lee & Magill, 1983; Shea & Morgan, 1979). It is assumed that variability between consecutive problems interferes with immediate performance since students have to use a new problem-solving procedure each time they encounter a new task. This interference leads to higher processing demands and lower performance during the acquisition phase, but results in better long-term retention and transfer performance later on (van Merriënboer et al., 2002). In light of this literature, one might expect that higher learning gains in the interleaved condition become apparent only in the posttest data, but not during the acquisition phase, because they might be "masked" by impaired performance due to contextual interference. Although my data does not confirm that interleaved practice schedules result in lower performance, my overall findings are in line with the notion that performance measures are not suitable for detecting differences between practice schedules during the acquisition phase. Rather than investigating differences between directly observed behaviors, Bayesian Knowledge Tracing models "machine-learn" a latent variable, namely the probability that a student transitions from the unlearned state to the learned state. These learning rate estimates appear to be a more suitable metric to detect advantages of interleaved practice even during the acquisition phase. In other words, "naïve" methods such as performance during the acquisition phase are not suitable to detect differences in students' learning from different practice schedules. Bayesian Knowledge Tracing analyses allow detecting learning gains that may be too subtle to detect during the acquisition phase when relying on student performance only.

4.3.4 Discussion

Taken together, the results from the learning outcomes, the think-aloud study, and the Bayesian Knowledge Tracing analysis yield interesting insights that are both of theoretical and practical significance. The results provide some support for the hypothesis that interleaving graphical representations leads to better learning than blocking graphical representations (while moderately interleaving task types). The analysis of the learning outcomes shows a significant advantage of interleaved practice only on transfer of conceptual domain knowledge at the delayed posttest, and a marginally significant advantage of the increasingly interleaved condition on reproduction with area models. The advantage of the fully interleaved condition over the blocked condition was particularly true for students with low prior knowledge. Taken together, the results from the learning outcomes provides only weak evidence that interleaved practice with graphical representations leads to better learning than blocked practice.

Yet, the analysis of the tutor log data with Bayesian Knowledge Tracing provides further support for this interpretation. The results show that a model that includes practice schedules fit the data best. Furthermore, I was able to detect advantages of interleaved practice over blocked practice even during the acquisition phase, in spite of students' equal problem-solving performance across conditions. These findings demonstrate that practice schedules of representations are a significant predictor of students' learning. In other words, the analysis of the tutor log data based on Bayesian Knowledge Tracing replicates the effect found in the learning outcomes that interleaving graphical representations leads to better learning than blocking graphical representations.

The results further show that the advantages of interleaved practice cannot only be detected based on long-term retention and transfer assessments (de Croock et al., 1998), but also by "machine-learning" a latent variable from students' problem-solving behaviors. To the best of my knowledge, the Experiment 3 is the first to empirically establish advantages of interleaved practice over blocked practice using data from the acquisition phase. These findings demonstrate that methods of educational data mining provide unique opportunities to gain deeper insights into educational psychology questions in a way that is not possible using “naïve” methods of looking at performance data alone. Furthermore, the analysis of the tutor log data provides further support for the conclusion that graphical representations should be presented in an interleaved fashion, rather than in a blocked fashion.

In addition, the results from Experiment 3 demonstrate significant learning gains for students who worked with a version of the Fractions Tutor that supports learning with multiple graphical representations of fractions, but not for those students who worked with only a single graphical representation. The gains persist until at least one week after the study when students completed the delayed posttest. This finding illustrates the merit of including multiple graphical representations of fractions, presented in an interleaved schedule, on students’ learning: within only six hours of supplementary practice with the Fractions Tutor, students significantly improved their knowledge in an important topic in math that presents a stumbling block for many students. Furthermore, this finding replicates and extends the finding from Experiment 1 by showing that a multiple-representations version of the Fractions Tutor leads to better learning than a variety of single-representation versions of the Fractions Tutor (and not just compared to a number-line only version, as used in Experiment 1).

Taken together, the findings from Experiment 3 extend prior research on interleaved practice schedules, which has focused on practice schedules of different task types. Experiment 3 demonstrates that the advantage of interleaved practice schedules of task types generalizes to practice schedules of multiple graphical representations, at least when it comes to supporting robust learning of conceptual domain knowledge. Although more research is needed to investigate whether these findings generalize to other domains, based on these findings I carefully conclude that designers of intelligent tutoring systems should employ an interleaved practice schedule of graphical representations in order to enhance robust conceptual knowledge of fractions. In providing specific design recommendations for how to use multiple graphical representations within instructional materials, Experiment 3 also extends research on learning with multiple representations: the findings show that practice schedules in which multiple representations are provided have an impact on their effectiveness (compared to a single representation). Since graphical representations are used across many STEM domains, this finding provides guidance for instructional designers of a wide range of instructional materials.

At first glance, it may seem that the findings from Experiment 2 and Experiment 3 contradict one another. In Experiment 2, blocking graphical representations (while interleaving task types) leads to better learning outcomes than interleaving graphical representations (while blocking task types) on a test of representational knowledge, which assessed conceptual understanding of graphical representations and the ability to transfer that knowledge to novel tasks. By contrast, in Experiment 3, interleaving graphical representations (while moderately interleaving task types) leads to better learning outcomes than blocking graphical representations (while moderately interleaving task types) on a test of reproduction with area models and on a test of transfer of conceptual knowledge, which included test items with and without graphical representations. Interpreting these seemingly contradicting findings is complicated by the fact that the topics covered by the Fractions Tutor in Experiment 2 and 3 were not the same, and that consequently, the tests used in Experiments 2 and 3 were not the same, since they reflect changes made to the Fractions Tutor between the two experiments.

Yet, after careful inspection, it becomes apparent that Experiments 2 and 3 make fundamentally different claims that do not necessarily contradict one another. Experiment 2 shows that interleaving task types has a *larger* impact on students' robust learning of representational knowledge than interleaving graphical representations, because task types are the more variable dimension on which learning tasks can differ, compared to graphical representations. In interpreting the findings from Experiment 2, I argued that blocking graphical representations might help students understand individual representations, as supported by the fact that the advantage of interleaving task types and blocking graphical representations (compared to interleaving graphical representations while blocking task types) was found on the representational knowledge scale used in Experiment 2.

By contrast, Experiment 3 investigates only the effect of interleaving graphical representations, while using a constant schedule of task types across all conditions. Rather than comparing the relative effect of interleaving one of two dimensions (i.e., task types and graphical representations) as in Experiment 2, Experiment 3 focuses on only one dimension: graphical representations. Given that Experiment 2 established that the effects of interleaving task types is *stronger* than that of interleaving graphical representations, it is not unexpected that the effect sizes found in Experiment 3 are relatively small. Yet, interleaving graphical representations may be *another* way in which repeated reactivation and abstraction across learning tasks can be supported. Interleaving multiple graphical representations may allow for deeper processing of the learning content as it provides more opportunities for students to reactivate representation-specific knowledge and for connection making between the different representations. The results from Experiment 3 support this conjecture: interleaving graphical representations enhances the acquisition robust domain knowledge as assessed by the conceptual transfer scale of the test used in Experiment 3. This finding might also shed further light into the results from Experiment 2. One caveat of Experiment 2 was that it could not tease apart whether the advantage of the int-types condition over the int-reps condition is due to the fact that task types were interleaved or that graphical representations were blocked. Although only a repetition of Experiment 3 that consistently uses a fully interleaved schedule of task types across all conditions could conclusively answer that question, the results from Experiment 3 suggest that the advantage of interleaving task types while blocking representations in Experiment 2 is due to the fact that task types were interleaved, and not that graphical representations were blocked.

Although the analysis of learning processes is based on the findings from only a small-scale think-aloud study conducted with only the fully interleaved condition, I believe that it provides interesting insights into the processes that underlie the benefits of interleaving multiple graphical representations. The results from the think-aloud study lead to the conclusion that interleaved practice does not appear to promote students' active connection making between different graphical representations across consecutive tutor problems. Rather, repeated reactivation of fractions knowledge that is specific to the graphical representation at hand appears to be the mechanism underlying the advantage of interleaved practice. When working with interleaved graphical representations, students have to reactivate the knowledge relevant to using that graphical representation to solve fractions problems more often than when working with blocked practice schedules of graphical representations. The process of loading representation-specific knowledge from long-term memory into working memory increases the strength of the association between the graphical representation and that knowledge, which in turn improves the likelihood that a student will be able to retrieve that knowledge later on. Future work should investigate whether indeed the advantage of interleaving representations results from repeated reactivation of knowledge about specific representations.

In interpreting the relation between Experiments 2 and 3, several open questions remain. First, as discussed above, the instructional materials and tests used in both experiments were not directly comparable. It would be interesting to repeat Experiment 2 with the same materials used in Experiment 3 to investigate whether the findings can be replicated. Second, Experiment 3 employed a moderately interleaved sequence of task types across all conditions, whereas the most successful condition in Experiment 2 used a highly interleaved sequence of task types. Therefore, Experiment 3 cannot answer the question whether the same advantage of interleaved practice applies to a scenario in which task types are also fully interleaved. It would be interesting to investigate whether the effects of interleaving task types and interleaving graphical representations interact. Interleaving both dimensions might have additive effects on students' learning, so that interleaving both dimensions is better than interleaving only one of them, or they might provide too much variation across task types so that students are overwhelmed and can no longer abstract common knowledge across the different learning tasks. Finally, it would be interesting to investigate the effects of interleaving graphical representations (and / or task types) in other domains than fractions learning. What type of knowledge can be abstracted across task types and graphical representations may differ across domains. However, to the extent that conceptual aspects of the domain knowledge occur across graphical representations and task types, I expect that the advantage of interleaved practice will generalize to other domains than fractions.

With respect to my theoretical framework for learning with multiple graphical representations, the mechanism that interleaving graphical representations supports may be representational fluency-building processes. Repeated reactivation of representation-specific knowledge may support the ease with which students can retrieve knowledge about individual graphical representations. The analysis of the tutor log data using Bayesian Knowledge Tracing provides some support for this interpretation: higher learning rates for the interleaved condition (compared to the blocked condition) indicate that students become more accurate at using individual graphical representations to solve fractions problems. This finding is what one would expect if students acquire representational fluency. Furthermore, the results from the think-aloud study suggest that repeated reactivation, rather than conscious abstraction across graphical representations, might be the mechanism that accounts for the advantage of interleaved over blocked practice with graphical representations. Reactivation might be more conducive to fluency-building processes than to sense-making processes. However, given the speculative nature of this interpretation, more research is needed to evaluate the claim that interleaving graphical representations promotes representational fluency. One caveat is that Experiment 3 did not include a direct test of representational fluency. Furthermore, the analysis of the tutor log data did not take latency measures of student response times while they worked with the Fractions Tutor into account. If students acquire fluency in using individual graphical representations, they should also become faster and more efficient at using them. To put the interpretation that interleaving graphical representations enhances students' representational fluency, which in turn promotes their learning of conceptual fractions knowledge, to a true test, one would have to assess representational fluency. My theoretical framework would predict that the hypothesized differences between interleaved and blocked practice on a representational fluency test explain the expected advantage of the interleaved over the blocked condition on the conceptual knowledge test.

Yet, this interpretation of the findings from Experiment 3 provide an interesting perspective also to reconcile the results from Experiments 2 and 3. As argued in Experiment 2, interleaving task types while blocking graphical representations may serve the development of *representational understanding*, by allowing students to gain in-depth understanding of how a given graphical representation applies across different task types. I argued that interleaving graphical representations may serve the development of *representational fluency*, as the repeated reactivation of representation-specific knowledge strengthens that knowledge and increases the chance that students can retrieve it fast and effortlessly later on. Thus, interleaving task types and graphical representations may serve complementary goals: interleaving task types while blocking graphical representations may enhance representational sense-making processes, whereas interleaving graphical representations (while moderately interleaving task types) may enhance representational fluency-building processes. At different times during the learning process, different relative sequences of multiple graphical representations 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 multiple graphical representations, speaks to this hypothesis. If representational fluency builds on representational understanding, blocking multiple graphical representations while interleaving task types might be most beneficial early in the learning sequence, whereas interleaving multiple graphical representations (in addition to interleaving task types) might become more important later during the learning process. Thus, if interleaving task types (while blocking graphical representations) supports representational sense-making processes, if interleaving graphical representations supports representational fluency-building processes, and if representational fluency builds on representational understanding, one might expect the best learning gains if different sequences are employed at different times during the learning process (or in particular, if the sequence of task types and graphical representations adapts to the level of representational understanding and representational fluency of each individual student).

However, as argued in Experiment 2, it is also possible that interleaving task types also supports representational fluency-building processes. It is indeed possible that both interleaving task types and interleaving graphical representations support the same learning process, by enhancing reactivation and abstraction across different aspects of fractions knowledge. If this is the case, and if reactivation is the mechanism that accounts for the advantage of interleaved practice, one might expect that fully interleaving both task types and graphical representations will lead to the best learning gains, as the higher variability of subsequent task types increases the amount of knowledge that students need to reactivate as they move between different task types and graphical representations. If, however, abstraction is the mechanism by which interleaved practice enhances learning, one might expect that interleaving both task types and graphical representations creates too much variability to abstract across subsequent learning tasks. In this case, further research is needed to compare varying degrees of interleaving task types and graphical representations, and to explore whether a potential “optimal” degree of variability between learning tasks differs between learners (e.g., low versus high prior knowledge students).

As the excessive use of if-clauses in the preceding two paragraphs illustrate, these interpretations are highly speculative and remain to be empirically tested in future research. The next steps in such research should be to identify the effects of practice schedules of task types and graphical representations on representational understanding and representational fluency directly (and separately from domain knowledge), and to identify the processes of reactivation and abstraction for the different types of interventions. Then, specific predictions about the relation between these different measures should be tested with appropriate mediation models.

Despite of its highly speculative nature, this interpretation of the findings from Experiment 2 and 3 illustrate the merit of my theoretical framework in generating new hypotheses that can be tested empirically based on the analysis of learning outcomes in experimental studies and formally evaluated through the use of data mining techniques on learning process measures. The suggested questions for future research would not only provide insights based on which the theoretical framework could be revised, but it would also provide further guidance for designers of educational materials that vary across task types and graphical representations, as is typically he case in many STEM domains.

4.4 Experiment 4: Combining connectional sense-making and fluency-building support

Experiments 1-3 focused on learning with multiple graphical representations that were provided across consecutive tutor problems, so that in each problem, only one graphical representation was present. Yet, prior research on multiple representations shows that students' benefit from multiple representations depends on their ability to make connections between them (Ainsworth, 2006; Cook et al., 2007; Even, 1998; Gutwill et al., 1999; Özgün-Koca, 2008; Plötzner et al., 2001; Plötzner et al., 2008; Schnotz & Bannert, 2003; Schwonke et al., 2008; Schwonke & Renkl, 2010; Taber, 2001). In the domain of fractions, connection making between different representations is considered to be an important learning goal (Cramer, 2001; Miura & Yamagishi, 2002; Taber, 2001). However, connection making is a difficult task for students that typically does not happen spontaneously (Ainsworth et al., 2002), especially when students have low prior knowledge (Bodemer & Faust, 2006). The results from the think-aloud study in Experiment 3 illustrate that students tend not to spontaneously connect consecutively presented graphical representations to one another, unless explicitly prompted to do so. The goal of Experiment 4 was to investigate the complementary role of instructional support for two learning processes hypothesized to be involved in connection making between multiple graphical representations: *connectional sense-making processes* and *connectional fluency-building processes*. This research question was motivated by my theoretical framework for learning with multiple graphical representations (see section 2.2). I expect that supporting connectional sense-making processes and connectional fluency-building should enhance students' benefit from multiple graphical representations. Yet, most research on connection making has focused on the support of either process, rather than on supporting both.

Research on connectional sense-making processes has typically investigated ways to help students relate representations based on 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 show that connectional sense-making processes are crucial for students' acquisition of domain knowledge. However, this research was exclusively conducted on supporting students in making connections between representations of different symbol systems: either between text and an additional graphical representation (Bodemer & Faust, 2006; Bodemer et al., 2004; Plötzner et al., 2001), or between symbolic representations and a corresponding graphical representation (van der Meij & de Jong, 2006). Whether these findings generalize to connection making between representations using the same symbol system (i.e., multiple graphical representations) remains an open question.

Another question that the literature on connectional sense-making support leaves open is the role of automated support provided by the system. On the one hand, providing students with auto-linked graphical representations (i.e., graphical representations that are linked in such a way that the student's manipulations of one are automatically reflected in the other) promotes learning in complex domains (van der Meij & de Jong, 2006). On the other hand, research shows that students should actively create connections between representations, rather than passively observing correspondences (Bodemer & Faust, 2006; Bodemer et al., 2004; Gutwill et al., 1999).

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-representation worked examples to single-representation worked examples and found that multiple representations can enhance students' learning from worked examples. Other studies (e.g., Berthold et al., 2008; Schwonke, Berthold et al., 2009) have also established that multiple representations can promote students' learning from worked examples. However, the question remains open whether worked examples can also enhance students' learning from multiple graphical representations.

Research on connectional fluency-building processes has 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 in Kellman and colleagues' studies learned to find corresponding representations of math problems, such as textual descriptions, graphical representations, and symbolic representations. Fluency training aims at helping students gain perceptual experience in making connections without asking them to consciously reflect on these connections. Rather, the training helps students become 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 good conceptual understanding of the domain and of the representations tend to score low on perceptual 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 such fluency tests, but also on tests of conceptual and procedural knowledge, compared to students who did not receive such training (Kellman et al., 2009). However, previous research on fluency-building processes in making connections has also taken a symbol-systems approach and only focused on connection making between representations using different symbol systems, rather than on connection making between representations using the same symbol system. It thus remains an open question whether Kellman and colleague's findings generalize to support of connectional fluency-building processes that involve multiple graphical representations.

4.4.1 Research questions and hypotheses

As argued, both connectional sense-making processes and fluency-building processes may play a role in students' learning with multiple graphical representations. Thus, Experiment 4 investigates the hypothesis that students will learn best about fractions when they receive support for both connectional sense-making processes and connectional fluency-building processes.

Furthermore, it may matter whether students are supported in actively engaging in connectional sense-making processes or whether they receive automated support to engage in connectional sense-making processes provided semi-automatically by the Fractions Tutor. A further goal of Experiment 4 is therefore to investigate whether students should become active in generating connections or receive automated support for connectional sense-making processes from the system.

Specifically, Experiment 4 investigated the following hypotheses:

Hypothesis 1: Support for connectional sense-making processes enhances students' acquisition of robust knowledge of fractions.

Hypothesis 2: Support for connectional fluency-building processes enhances students' acquisition of robust knowledge of fractions.

Hypothesis 3: Support for both connectional sense-making and fluency-building processes enhances students' acquisition of robust knowledge of fractions more so than either type of support alone.

Hypothesis 4: Students' acquisition of robust knowledge of fractions benefits more from a type of connectional sense-making support that requires them to actively make sense of connections themselves, rather than making sense of correspondences that are depicted semi-automatically.

4.4.2 Methods

To investigate these hypotheses, I conducted a classroom experiment that contrasted the effects of different types of support for connection making between multiple graphical representations of fractions.

4.4.2.1. Experimental design

Students were randomly assigned to one of seven experimental conditions, summarized in Table 18. There were two different types of connection-making support: sense-making support and fluency-building support. Students were assigned to either receiving sense-making support for connection making (in the form of either auto-linked support or worked-example support) or not. This factor was crossed with a second experimental factor, namely, whether or not students received fluency-building 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. In other words, two conditions did not receive any type of connection-making support: the number-line condition and the multiple graphical representations condition.

		Sense-making support			**Control**
		None	***Auto-linked representations***	***Worked example***	
Fluency-building support	***None***	Multiple graphical representations condition	Auto-linked condition	Worked-example condition	
	Fluency-building support	Fluency-building condition	Auto-linking and fluency-building condition	Worked-example and fluency-building condition	
Control					Number-line condition

Table 18. Experiment 4: Overview of experimental conditions.

4.4.2.2. Participants

A total of 1308 4th- and 5th-grade students from three school districts (13 schools, 62 classes) participated in the experiment during their regular math class. Due to technical failure that resulted in data loss, only one school district (5 schools, 25 classes, N = 599) had complete data. I excluded students who did not complete all tests and who did not complete their work on the tutoring system, yielding a total of N = 428. The number of students who were excluded from the analysis did not differ between conditions, χ^2 (6, N = 169) = 4.34, p > .10. Table 19 shows the number of included and excluded students per condition.

	included	excluded
Number-line condition	61	24
Multiple graphical representations condition	64	30
Auto-linked condition	52	25
Worked-example condition	59	28
Fluency-building condition	73	18
Auto-linking and fluency-building condition	61	24
Worked-example and fluency-building condition	58	20

Table 19. Experiment 4: Number of students included and excluded by condition.

4.4.2.3. Procedure

The study took place at the beginning of the 2011/2012 school year. Students' regular math teachers led the sessions, but researchers were present in the classrooms during the first two days of the study to assist teachers in answering questions specific to the use of the Fractions Tutor.

Students accessed all materials online from their school's computer lab. On the first day of the study, students completed a 30-minute pretest. They then worked on the Fractions Tutor for about ten hours, spread across consecutive school days (with the version described in section 3.4, on topics 1-10 in Table 1, see section 3.4). The day following the tutor sessions, students completed a 30-minute posttest and took a 5-minute survey. About one week after the posttest, students were given an equivalent delayed posttest.

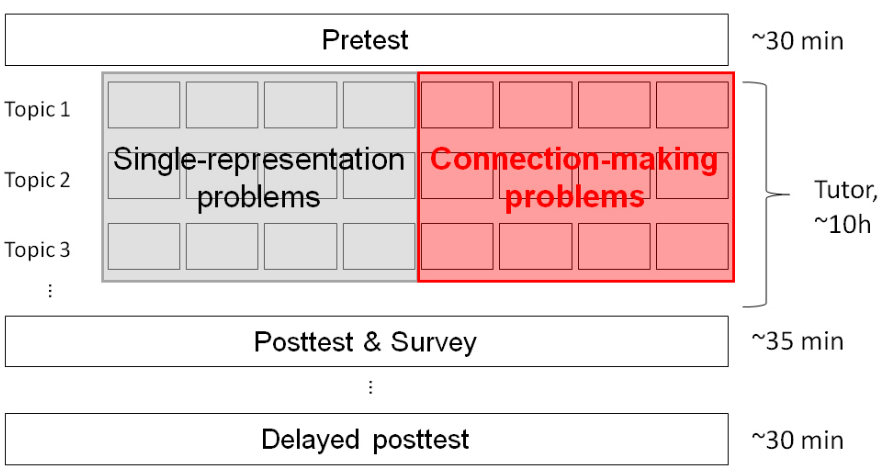


Fig. 24. Experiment 4: Procedure for conditions with connection-making support.

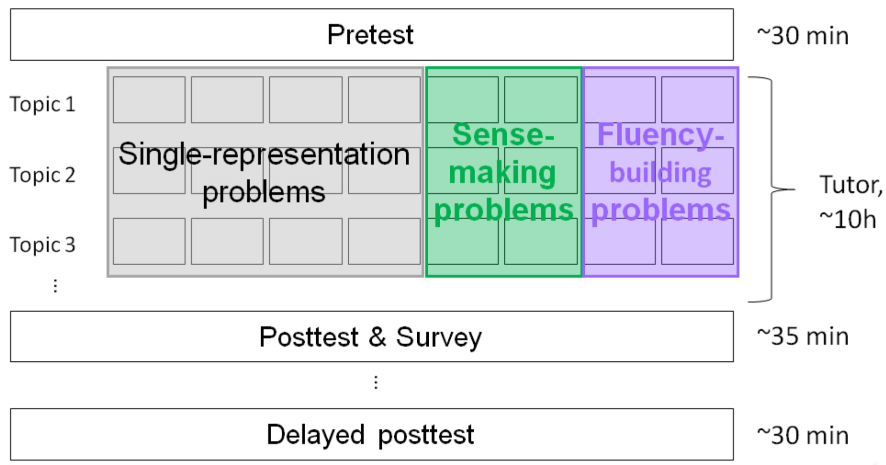


Fig. 25. Experiment 4: Procedure for conditions with sense-making and fluency-building support.

Fig. 24 illustrates this procedure for students in all of the conditions with connection-making support (i.e., in all but the multiple graphical representations condition and the number-line condition, see Table 18). For each of the topics covered by the Fractions Tutor, students first completed four single-representation tutor problems, followed by four connection-making problems. The four connection-making problems were assigned according to the student's condition. That is, students in the auto-linked condition received four auto-linked problems, students in the worked-example condition received four worked-example problems, and students in the fluency-building condition received four fluency-building problems. Students in the conditions with sense-making *and* fluency-building support received two problems of each kind, as illustrated in Fig. 25. That is, students in the auto-linking and fluency-building condition received two auto-linked problems followed by two fluency-building problems, and students in the worked-example and fluency-building condition received two worked-example problems followed by two fluency-building problems. Students in the conditions without connection-making support (i.e., the multiple graphical representations condition and the number-line condition) received eight single-representation problems per topic. All tutor problems involved a comparable number of problem-solving steps and took about the same time to complete.

4.4.2.4. Fractions Tutor versions

Students in Experiment 4 worked with the latest version of the Fractions Tutor on all ten topics. The single-representations problems, worked-example problems, and fluency-building problems were described in detail in sections 3.4.2 and 3.4.3, respectively.

Fig. 26 shows an example of an auto-linked problem. The auto-linked problems followed the same side-by-side format as the worked-example problems, but there were no worked examples. Rather, students interacted with a number line to solve a problem, while an area model representation (i.e., a circle or a rectangle) updated automatically to mimic the steps the student performed on the number line. In this sense, the more familiar representation provided feedback on the work with the less familiar representation. (At a technical level, the number line CTAT component [Aleven et al., 2009] served as a controller for the circle and rectangle CTAT components.) As the worked-example problems, the auto-linked problems included reflection prompts at the end of each problem (see bottom of Fig. 26), which asked students to identify correspondences of the two given representations. The inclusion of reflection prompts in the sense-making problems was motivated both by the results from Experiment 1, and by the finding in the think-aloud study in Experiment 3 that students can generate connections between graphical representations when prompted to do so.

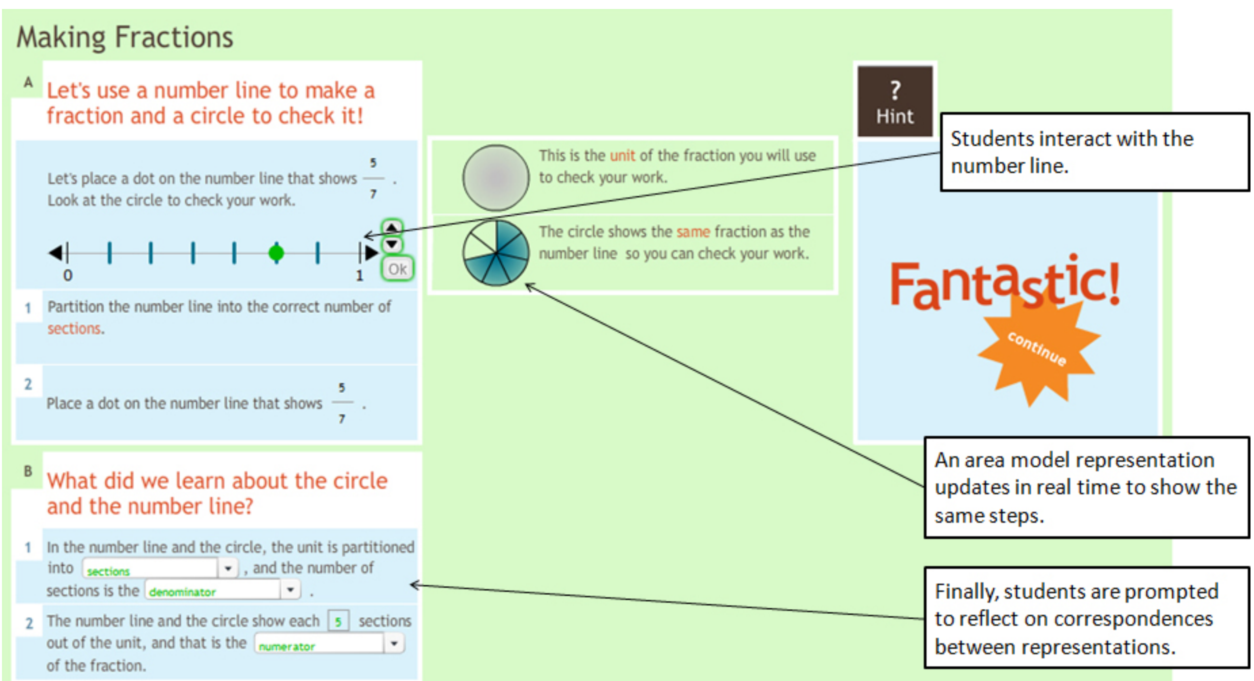


Fig. 26. Experiment 4: Example of an auto-linked problem to support connectional sense-making processes.

4.4.2.5. Test instruments

Students took three tests: a pretest, an immediate posttest, and a delayed posttest. The tests were adapted from the tests used in Experiment 3, but included a variety of new test items to reflect changes made to the Fractions Tutor curriculum. I created three equivalent test forms, which included the same type of problems, but with different numbers. Based on data from a pilot study with 61 4th-grade students, I made sure that the difficulty level of the test was appropriate for the target age group, and that the different test forms did not differ in difficulty. For the classroom study, I randomized the order in which the different test forms were administered.

The tests targeted two knowledge types: procedural and conceptual knowledge. The conceptual knowledge scale assessed students' principled understanding of fractions; it assessed similar aspects of conceptual knowledge of fractions as the conceptual transfer scale in Experiment 3, and it also includes items similar to the representational knowledge scale used in Experiment 2, but in a way that corresponds to the topics covered in the revised version of the Fractions Tutor (described below). Examples of the test items can be found in Appendix 7. The test items included reconstructing the unit, identifying fractions from graphical representations, proportional reasoning questions, and verbal reasoning questions about comparison tasks. The procedural knowledge scale assessed students' ability to solve questions by applying algorithms; it assessed similar aspects of procedural knowledge as the operational knowledge scale used in Experiment 2 and the procedural transfer scale used in Experiment 3, but in a way that aligns with the topics covered in the revised version of the Fractions Tutor. The test items included finding a fraction between two given fractions using representations, finding equivalent fractions, addition, and subtraction. The theoretical structure of the test (i.e., the two knowledge types just mentioned) was based on a factor analysis with the pretest data from the current experiment. Half of the items in both test scales were reproduction and transfer items, respectively. I validated the resulting factor structure using the data from the immediate and the delayed posttests.

4.4.3 Results

		pretest	immediate posttest	delayed posttest
conceptual knowledge	Multiple graphical representations	.33 (.20)	.45 (.23)	.48 (.26)
	Auto-linking	.38 (.20)	.49 (.23)	.51 (.26)
	Worked examples	.36 (.22)	.43 (.20)	.49 (.26)
	Fluency-building	.31 (.21)	.37 (.22)	.44 (.24)
	Auto-linking and fluency-building	.36 (.20)	.43 (.24)	.49 (.25)
	Worked examples and fluency-building	.39 (.21)	.52 (.24)	.58 (.26)
	Number-line only	.37 (.20)	.43 (.25)	.48 (.20)
procedural knowledge	Multiple graphical representations	.25 (.25)	.30 (.28)	.30 (.26)
	Auto-linking	.21 (.18)	.26 (.24)	.26 (.24)
	Worked examples	.26 (.21)	.29 (.24)	.31 (.27)
	Fluency-building	.19 (.17)	.23 (.20)	.25 (.22)
	Auto-linking and fluency-building	.20 (.18)	.25 (.21)	.26 (.21)
	Worked examples and fluency-building	.26 (.20)	.32 (.26)	.33 (.26)
	Number-line only	.21 (.20)	.25 (.22)	.27 (.23)

Table 20. Experiment 4: Means and standard deviations (in parentheses) for conceptual and procedural knowledge at pretest, immediate posttest, delayed posttest. Min. score is 0, max. score is 1.

measure	effect	significant	F/t-value	adj. *p*-value	effect size
conceptual knowledge	sense-making support	no	$F < 1$		
	fluency-building support	no	$F < 1$		
	sense-making * fluency-building support	yes	$F(2, 351) = 3.97$	$p < .05$	$\eta^2 = .03$
	worked-example and fluency-building > number-line control	yes	$t(115) = 2.41$	$p < .05$	$d = .27$
	effect slices for the effect of sense-making support	yes	$F(2, 343) = 4.34$	$p < .05$	$\eta^2 = .07$
	worked-example and fluency-building > auto-linking and fluency-building	yes	$t(342) = 2.20$	$p < .05$	$d = .26$
	worked-example and fluency-building > fluency-building	yes	$t(341) = 2.82$	$p < .01$	$d = .32$
procedural knowledge	sense-making support	no	$F < 1$		
	fluency-building support	no	$F < 1$		
	sense-making * fluency-building support	no	$F < 1$		
	effect slices for the effect of sense-making support	no	$F < 1$		

Table 21. Experiment 4: Results on conceptual and procedural knowledge.

As mentioned, I analyzed the data of $N = 428$ students. Table 20 shows the means and standard deviations for the conceptual and procedural knowledge scales by test time and condition. Table 21 gives an overview of the results.

4.4.3.1. Effects of connection-making support on learning outcomes

I used an HLM (see Raudenbush & Bryk, 2002) with four nested levels to analyze the data. At level 1, I modeled performance on each of the tests for each student. At level 2, I accounted for differences between students. Level 3 models random differences between classes, and level 4 random differences between schools. The HLM is the outcome of a forwards-inclusion procedure in which I used the Bayesian Information Criterion (BIC) to find whether the inclusion of a variable increased model fit. A lower BIC indicates better model fit while penalizing for greater model complexity. If the BIC decreased as a consequence of including a variable, I kept the variable. If the BIC did not decrease, I did not include the variable. I 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. More specifically, the following HLM was fitted to the data:

$$Y_{ijkl} = (((\mu + W_l) + V_{kl}) + \beta_2 * s_j + \beta_3 * f_j + \beta_4 * p_j + \beta_5 * s_j * p_j + \beta_6 * f_j * p_j + U_{jkl}) + \beta_1 * t_i + R_{ijkl} \quad (6)$$

with

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

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

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

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

with the index *i* standing 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 support s_j, β_3 is the parameter for the effect of fluency-building support f_j, β_4 is the parameter for the effect of student$_j$'s performance on the pretest p_j, β_5 is the parameter for an aptitude-treatment interaction between sense-making support s_j and student$_j$'s performance on pretest p_j, β_6 is the parameter for an aptitude-treatment interaction between fluency-building support 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.

Since the HLM described in (6) uses students' pretest scores as a covariate, it does not allow us to analyze whether students in the various conditions improved from pretest to immediate and delayed posttest. To analyze learning gains, I excluded pretest score in the dependent variable, as well as an interactions of pretest with sense-making support and fluency-building support, yielding:

$$Y_{ijkl} = (((\mu + W_l) + V_{kl}) + \beta_2 * s_j + \beta_3 * f_j + U_{jkl}) + \beta_1 * t_i + R_{ijkl} \quad (7)$$

with

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

(level 2) $\varepsilon_{jkl} = \delta_{kl} + \beta_2 * s_j + \beta_3 * f_j + U_{jkl}$

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

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

with the index *i* standing for test time (i.e., pretest, immediate, and delayed posttest). The dependent variable Y_{ijkl} is student_j's score on the dependent measures at test time t_i (i.e., pretest, immediate posttest, or delayed posttest). Excluded were the parameters β_4 for the effect of student_j's performance on the pretest p_j, parameter β_5 is for an aptitude-treatment interaction between sense-making support s_j and student_j's performance on pretest p_j, and parameter β_5 is for an aptitude-treatment interaction between fluency-building support f_j and student_j's performance on pretest p_j.

In addition, I specified posthoc comparisons within the HLM to clarify the effects of sense-making support and fluency-building support. The reported *p*-values were adjusted for multiple comparisons using the Bonferroni correction. I report partial η^2 for effect sizes on main effects and interactions between factors, and Cohen's *d* for effect sizes of pairwise comparisons. According to common convention (Cohen, 1988), an effect size partial η^2 of .01 is considered a small effect, .06 a medium effect, and .14 a large effect. An effect size *d* of .20 corresponds to a small effect, .50 to a medium effect, and .80 to a large effect.

Effects of sense-making support and fluency-building support

To investigate hypothesis 1 (that support for connectional sense-making processes enhances students' acquisition of fractions knowledge), I computed the main effect of sense-making support using the hierarchical linear model described in equation 6. There was no main effect of sense-making support on conceptual knowledge ($F < 1$) nor on procedural knowledge ($F < 1$).

To investigate hypothesis 2 (that support for connectional fluency-building processes enhances students' acquisition of fractions knowledge), I computed the main effect of fluency-building support. There was no main effect of fluency-building support on conceptual knowledge ($F < 1$) nor on procedural knowledge ($F < 1$).

To investigate hypothesis 3 (that students' acquisition of fractions knowledge requires support for both connectional sense-making and fluency-building processes), I computed the interaction effect between sense-making support and fluency-building support. The results provide support for hypothesis 3: there was a significant interaction effect between sense-making and fluency support on conceptual knowledge, $F(2, 351) = 3.97$, $p < .05$, $\eta^2 = .03$, such that students who received both types of support performed best on the conceptual knowledge posttests. There was no significant interaction effect on procedural knowledge ($F < 1$).

Finally, to verify the advantage of receiving connection-making support over the number line control condition, I compared the most successful condition (worked-example and fluency-building) to the number line control condition using *posthoc* comparisons. The advantage of the worked-example and fluency-building condition over the control condition was significant on conceptual knowledge, $t(115) = 2.41$, $p < .05$, $d = .27$.

In summary, the interaction effect between sense-making support and fluency-building support shows that supporting both sense-making processes and fluency-building processes in connection making enhance students' conceptual knowledge of fractions more so than either type of support alone. Since neither type of support alone enhanced students' learning of fractions, and since only the condition that received a combination of worked-example support for sense-making processes and fluency-building support outperformed the number-line only condition, I conclude that both types of support *together* are *necessary* for students to benefit from multiple graphical representations.

Effects of active engagement in sense-making processes

To investigate hypothesis 4 (that students' acquisition of fractions knowledge benefits more from a type of connectional sense-making support that requires students to actively make sense of connections), I computed effect slices using the hierarchical linear model described in equation 6: a test of the effect of sense-making support for each level of the fluency support factor. Effect slices are a way of performing a partitioned analysis of one experimental for a given level of a second experimental factor. I computed effect slices for sense-making support (i.e., a test of the effect of sense-making support) for each level of the fluency-building support factor within the

HLM shown described in equation 6. The test of effect slices for sense-making support (i.e., a test of the effect of sense-making support for each level of the fluency support factor) showed that there was a significant effect of sense-making support within the conditions with fluency support on conceptual knowledge, $F(2, 343) = 4.34$, $p < .05$, $\eta^2 = .07$, but not within the conditions without fluency support ($F < 1$). *Posthoc* comparisons between the fluency-building condition, auto-linking and fluency-building condition, and the worked-example and fluency-building conditions confirmed that the worked-example and fluency-building condition significantly outperformed the fluency-building condition, $t(341) = 2.82$, $p < .01$, $d = .32$, and the auto-linking and fluency-building condition $t(342) = 2.20$, $p < .05$, $d = .26$, on conceptual knowledge. In summary, worked-example problems are more effective in supporting sense-making of connections than auto-linked problems, provided that students also receive fluency support.

Taken together, Experiment 4 provides evidence that the combination of both connectional sense-making and fluency-building support enhances students' robust learning of conceptual knowledge. Furthermore, the results demonstrate that students need to actively engage in connectional sense-making activities. A novel application of worked examples is a successful means to support students' connectional sense-making processes. Finally, the fact that only the combined worked-example and fluency-building support condition significantly outperformed the number-line control condition demonstrates that students' benefit from multiple graphical representations depends on receiving support for both connectional sense-making and fluency-building processes.

4.4.3.2. Causal path modeling

By integrating two different, thus far separate lines research on connectional sense-making and fluency-building processes, Experiment 4 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 conditions that demonstrated significant differences on the conceptual posttests, that is, in errors the worked-example condition, fluency-building condition, and combined worked-example and fluency-building con-

Error Type	**Description**	**# in worked-example condition**	**# in fluency-building condition**	**# in worked-example and fluency-building condition**
additionMixedError	Finding representations that show the addend of a given sum equation depicted by representations	n/a	207	176
compareMixed-Error[3]	Finding representations that show a fraction smaller or larger than the given one	n/a	436	307
comparisonError	Comparing two fractions	92	n/a	82
denomError[7]	Entering the denominator of a fraction	972	n/a	837
DiffMixedError	Finding representations that show the difference of two fractions	n/a	282	238
equivalence-CompareError	Judging whether two fractions are equivalent	19	n/a	18
equivalenceError[6,4]	*Finding equivalent fraction representations*	*n/a*	*2899*	*2157*
improperMixed-Error[6,7]	*Finding representations of improper fractions*	*n/a*	*1380*	*1608*
multiplyError	Entering a number by which to multiply numerator or denominator to expand a given fraction	30	n/a	29
nameCircleMixed-Error[6,7]	*Finding circle representations that show the same fraction as a number line or a rectangle*	*n/a*	*355*	*126*
nameNLMixed-Error[6]	Finding number line representations that show the same fraction as a circle or a rectangle	n/a	949	599
nameRectMixed-Error[6]	Finding rectangle representations that show the same fraction as a number line or a circle	n/a	385	133
nlPartitionError[6]	Partitioning the number line to show an equivalent fraction	1913	n/a	2115
numberSections-UnitError	Finding the denominator of a fraction by indicating how many sections the unit was divided into	41	n/a	44
numError[6]	Entering the numerator of a fraction	1559	n/a	1390
place1Error[6,7]	*Locating 1 on the number line given a dot on the num-*	*150*	*n/a*	*222*

[3] Significant difference between conditions, based on chi-square test
[4] Significant predictor of performance on conceptual posttest, after controlling for pretest performance

	ber line and the fraction it shows			
placeDotError[6]	Placing a dot on the number line to show a fraction	198	n/a	253
sectionsBetween-0-1	Indicating that the denominator in a number line is shown by the sections between 0 and 1	61	n/a	44
SE-Error[6,7]	*Self-explanation error, response to reflection questions*	*1320*	*n/a*	*1629*
subtractionMixed-Error[7]	Finding representations that show the subtrahend of a given difference equation depicted by representations	n/a	214	240
sumMixedError[6]	Finding representations that show the sum of two fractions	n/a	256	205
unitError	Selecting the unit for a fraction given the symbolic fraction and a graphical representation	123	n/a	115
unitMixedError	Finding the unit of a given fraction	n/a	1050	1138

Table 22. Experiment 4: Error types and number of error types per condition. *Error types in italics* were selected for further analysis.

ditions made on the worked example problems and on the fluency-building problems. Rather than using the overall error rate, I applied a knowledge component model that underlies the problem structure of the Fractions Tutor to the errors while working on the tutor problems. Doing so allows for a more fine-grained analysis of students' errors than the overall error rate does. 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 worked-example condition and the worked-example and fluency-building condition, and errors on fluency-building problems in the fluency-building condition and the worked-example and fluency-building condition. Table 22 summarizes the types of errors that were possible in worked-example problems and in fluency-building problems.

I included only those error types into further analysis, which (1) were significant predictors of students' posttest performance, while controlling for pretest performance, and (2) significantly differed between conditions. To determine whether an error type was a significant predictor of students' immediate posttest performance, I conducted linear regression analyses with posttest performance as the dependent variable, and pretest performance and number of error type as predictors. For both the chi-square tests and the regression analyses, I controlled for multiple comparisons using the Bonferroni correction.

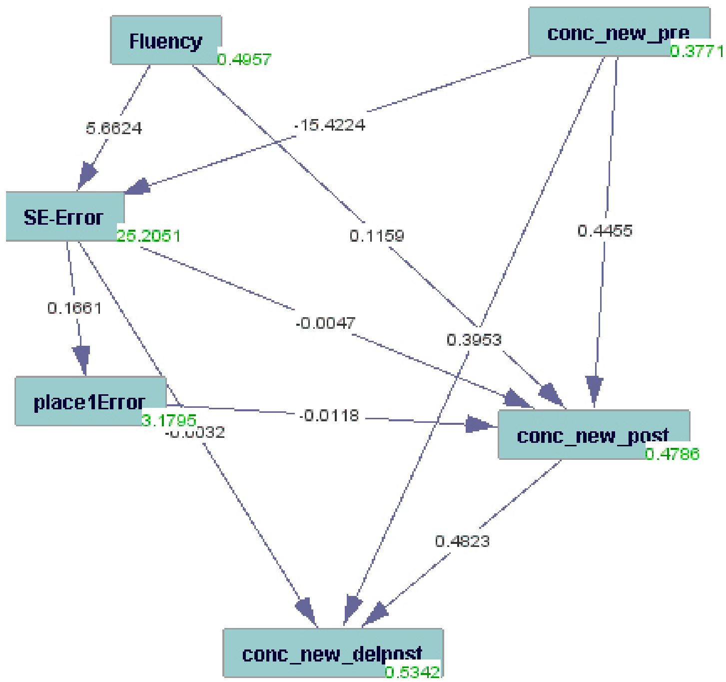


Fig. 27. Experiment 4: Structural equation model for the worked-example condition and the worked-example and fluency-building condition. Fluency: whether or not students received fluency-building support in addition to worked-example support; conc_new-pre: performance on conceptual pretest; conc_new_post: performance on conceptual immediate posttest; conc_new_delpost: performance on conceptual delayed posttest.

Fig. 27 shows the results from the structural equation model for the comparison of the worked-example support condition and the worked-example and fluency-building condition based on errors students made on the worked example problems. Students in the worked-example and fluency-building condition, compared to the worked-example 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 structural equation model for the worked-example condition and the worked-example and fluency-building condition 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. In general, since students who receive worked-example and fluency-builidng support work on fewer worked-example problems (2 per topic) than students who receive only worked-example support (4 per topic) are expected to make more errors on steps in the worked-example problems, simply because they receive less practice on them. Yet, the two particular error types on which I found differences between conditions, SE errors and place1 errors, on which fluency-building support had a negative effect, might correspond to steps in the tutor that particularly required conceptual understanding of connections. SE errors (i.e., errors in answering self-explanation prompts) directly require students to answer questions about *which aspects* of two graphical representations correspond to one another. Place1 errors (i.e., errors in finding 1 on a number line given a dot that shows a fraction), requires students to conceptually understand the concept of the unit of a fraction, which is a particularly hard concept to understand with number lines, but that is relatively easy to grasp with area models, such that sense-making support may have especially helped students understand this concept in particular. Taken together, the only mediation of the effect of the worked-example and fluency-building condition compared to the worked-example condition was negative, but there is no mediation of a positive effect, only a direct positive effect.

Fig. 28 shows the best-fitting structural equation model for the fluency-building condition and the worked-example and fluency-building support condition based on the errors students made on fluency-building problems. Students in the worked-example and fluency-building 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 fluency-building 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.

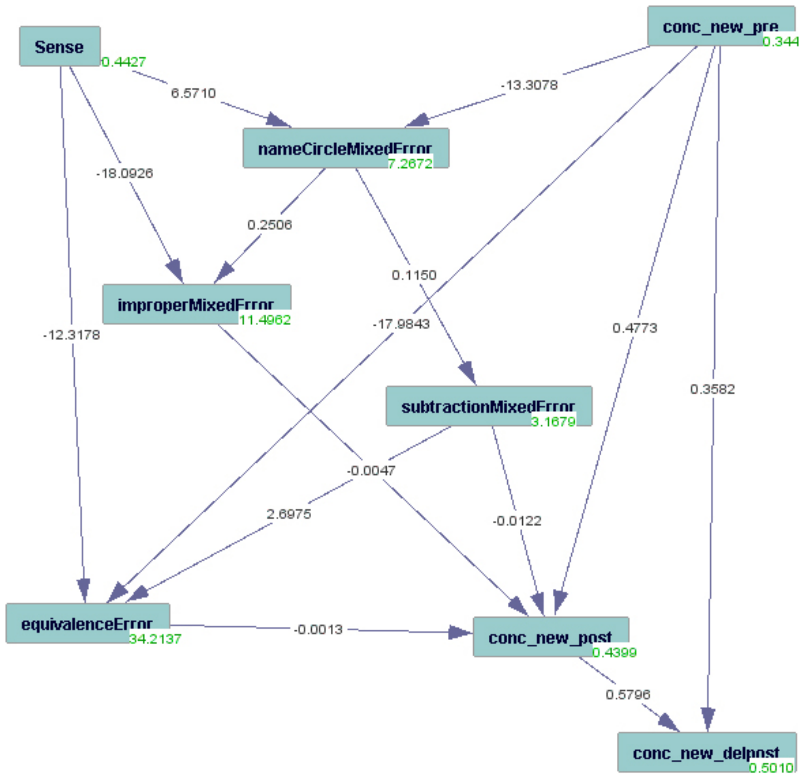


Fig. 28. Experiment 4: Structural equation model for the fluency-building condition and the worked-example and fluency-building condition. Sense: whether or not students received worked-example support in addition to fluency-building support; conc_new-pre: performance on conceptual pretest; conc_new_post: performance on conceptual immediate posttest; conc_new_delpost: performance on conceptual knowledge test.

The findings from the structural equation model 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 worked-example problems, which were associated with higher posttest performance. The structural equation model 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 structural equation model analysis does not support the notion that fluency-building support helps students benefit from sense-making problems.

4.4.4 Discussion

Experiment 4 tested a hypothesis that was deducted from my theoretical framework for learning with multiple graphical representations: that instructional support for connectional sense-making processes and support for fluency-building processes enhance students' learning of domain knowledge. The results from Experiment 4 provide support for this hypothesis only with respect to the combination of both types of support: taken alone, neither sense-making support nor fluency-building support were effective at enhancing students' learning of conceptual knowledge, but only together did they enhance students' learning. There were no effects of connection-making support on procedural knowledge. Furthermore, only students who received both types of instructional support for connection making significantly outperformed the number-line control condition on conceptual knowledge. With regard to the question of how to implement instructional support for connectional sense-making processes, Experiment 4 shows that a type of sense-making support that requires students to become active in generating connections, rather than observing connections provided by auto-linked representations, is the most effective.

Experiment 4 extends previous research on connection making that has (1) only focused on supporting either connectional sense-making processes or connectional fluency-building processes but not on supporting both types of learning processes, and (2) only focused on connection making between multiple representations of different symbol systems but not on connection making between multiple graphical representations.

Experiment 4 did not replicate earlier findings from Experiments 1 and 3, that multiple graphical representations provided without instructional support for connection making lead to better learning than a single graphical representation. This finding is surprising, especially since the multiple graphical representations condition corresponded directly to the fully interleaved condition in Experiment 3. It is possible that changes in the participant sample (different schools, different teachers), changes made to the Fractions Tutor (revised curriculum), or changes made to the test in consequence to the revisions of the Fractions Tutor may have affected the replicability of the advantage of multiple over a single graphical representation. In particular, the advantage of multiple graphical representations in Experiment 3 was on the conceptual transfer scale of the test, whereas the conceptual knowledge test used in Experiment 4 included both reproduction and transfer items. It is possible that the advantage of multiple graphical representations manifests itself most strongly in the ability to transfer conceptual domain knowledge to novel problems. Further research should investigate which knowledge types particularly benefit from multiple graphical representations. On the flip side, the results from Experiment 4 show that one might expect a stronger advantage of multiple graphical representations over a single graphical representation in Experiments 1 and 3 if explicit connection making support had been provided in these experiments as well.

With regard to my theoretical framework, Experiment 4 provides support for the claim that both connectional sense-making and fluency-building processes play a role in students' learning with multiple graphical representations. However, Experiment 4 did not assess connectional understanding and connectional fluency: instead, I infer from the fact that instructional support designed to enhance connectional sense-making processes and connectional fluency-building processes promotes students' learning of robust domain knowledge that these learning processes play a role in learning with multiple graphical representations. A direct test of this claim would also assess connectional understanding and connectional fluency and investigate whether differences on these measures can explain the differences in robust domain knowledge.

Taken together, Experiment 4 suggests that connectional sense-making processes and fluency-building processes interact. But, it remains unclear *how* these learning processes interact. Experiment 4 leads to the question whether both types of support simply need to be present, whether students' benefit from connectional sense-making support depends on their acquisition of connectional fluency, or whether students' benefit from connectional sense-making support depends on their acquisition of connectional understanding. The causal path analysis on the tutor log data obtained from Experiment 4 leads to the hypothesis that the acquisition of connectional understanding provides the foundation for students' benefit from support for connectional fluency-building processes. Experiment 5 was designed to investigate this hypothesis.

4.5 Experiment 5: Sequencing connectional sense-making and fluency-building support

Experiment 4 shows that support for both connectional sense-making processes and connectional fluency-building processes enhance students' benefit from multiple graphical representations. While it is evident that connectional sense-making support and connectional fluency-building support interact, it remains unclear *how* they interact. Does connectional understanding enhance students' benefit from fluency-building support, or does connectional fluency enhance their benefit from connectional sense-making support? This question arises from the findings in Experiment 4, in which neither connectional sense-making support nor connectional fluency-building support alone were effective, but only both types of support together enhanced students' learning of conceptual fractions knowledge. Based on a structural equation model of students' error types, one might hypothesize that connectional sense-making support helps students to benefit from connectional fluency-building support, rather than the other way around. But the structural equation model analysis cannot conclusively answer this question: in Experiment 4, connectional fluency-building support was consistently provided after connectional sense-making support. It is unclear whether connectional fluency-building support might have helped students benefit from connectional sense-making support if the two types of support had been provided in the reverse order. Further, due to the selective nature of the structural equation model analysis (i.e., selection of error types only if they significantly differed between conditions and predicted posttest performance, see section 4.4.3.2), its merit can only be hypothesis generation, but not empirical evidence for (or rather, against) a hypothesis.

4.5.1 Theoretical perspectives and hypotheses

The question of how connectional sense-making support and fluency-building support interact is also of practical relevance. If connectional understanding (acquired through support for connectional sense-making processes) enables students to benefit from connectional fluency-building support, instructional designers should provide support for connectional sense-making processes before support for connectional fluency-building processes (*understanding-first hypothesis*). If, on the other hand, connectional fluency (acquired through support for connectional fluency-building processes) enables students to benefit from connectional sense-making support, they should support connectional fluency-building processes before connectional 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' learning of robust domain knowledge.

By investigating how connectional sense-making support and fluency-building support interact, my research is a step towards closing the gap between studies that have exclusively focused on sense-making support (e.g., Bodemer & Faust, 2006; Bodemer et al., 2004; Brünken et al., 2005; Seufert & Brünken, 2006; van der Meij & de Jong, 2006) and those that have focused solely on fluency-building support (e.g., Kellman & Garrigan, 2009; Kellman et al., 2008; Kellman et al., 2009). Furthermore, Experiment 5 extends this prior research by focusing on multiple graphical representations that use the same symbol system. In this section, I describe theoretical perspectives pertaining to these competing hypotheses on the question of which learning process to support first.

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., 2012; Ohlsson, 2008) describe both sense-making processes and fluency-building processes as integral learning processes that students need to perform to master a domain. Furthermore, although many educational practice guides for math 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. Yet, 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. Research on fluency-building processes, however, has mostly focused on fluency in fact retrieval (Arroyo et al., 2011; Benjamin, Bjork, & Schwartz, 1998; Johnson & Layng, 1996), rather than on the type of *perceptual* fluency in making connections between graphical representations that I consider here, as described by Kellman and colleagues (Kellman & Garrigan, 2009; Kellman et al., 2008; Kellman et al., 2009). Unfortunately, neither of these literatures makes explicit claims about dependences between sense-making processes and fluency-building processes, which might have obviated the need for an empirical investigation of this question. In the following, I summarize arguments that speak for the hypothesis that one might expect the most robust learning gains when supporting connectional fluency-building processes before connectional sense-making processes (fluency-first hypothesis), and arguments for the opposite prediction, that instruction will be most effective when connectional sense-making processes are supported before connectional fluency-building processes (understanding-first hypothesis). I will then discuss specific predictions made by each of these hypotheses that can be tested empirically.

4.5.1.1. Fluency-first hypothesis

Students who acquire connectional fluency, that is, who are fluent in making connections between multiple graphical representations by visually relating them, may benefit from increased "cognitive head room" during subsequent learning tasks (Koedinger et al., 2012) that involve sense-making processes about the conceptual nature of the connections between multiple graphical representations.

Indeed, Kellman and colleagues (2009) argue that fluency results from automation of the perceptual task to make connections between different representations. This type of connectional fluency is acquired through experience with a variety of representations without having to engage in sense-making processes about how corresponding knowledge components are depicted in the different graphical representations. 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 that might be needed for the completion of a more complex task; these might well be tasks that involve connectional sense-making processes. If so, providing connectional fluency-building support before connectional sense-making support may decrease the risk of cognitive overload while students work on connectional sense-making problems, which is known to hamper learning (Chandler & Sweller, 1991).

4.5.1.2. Understanding-first hypothesis

Kellman et al.'s (2009) account relies on the assumption that connectional fluency – the automation of perceptually relating multiple graphical representations – *can* be learned independently from understanding of these connections. However, this assumption might not be true: connectional understanding might equip students with the knowledge they need in order to benefit from connectional fluency-building support. If students do not know what aspects of different graphical representations correspond to one another, how should they know what to attend to while working on connectional 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 math 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 (e.g., "In grades 3-5 all students should – develop understanding of fractions as parts of unit wholes, as parts of a collection, as locations on number lines, and as divisions of whole numbers", from NCTM Number and Operation Standards). The ability to efficiently work with fractions representations is expected later – not before the end of grade 8 (e.g., "In grades 6–8 all students should – work flexibly with fractions, decimals, and percents to solve problems", from NCTM Number and Operation Standards).

While Bieda and Nathan (2009) discuss how fluency in connection making between algebra representations may help students acquire abstract understanding of algebraic concepts and to transfer that knowledge to novel tasks, they also describe the danger of students being overly influenced by the visual properties of a representation, rather than by conceptual understanding of the representation. Thus, although they did not experimentally investigate this assertion, they suggest that connectional understanding may enable students to pay attention to the conceptually relevant aspects of the representations when developing connectional fluency.

4.5.1.3. Specific predictions

Both the understanding-first hypotheses and the fluency-first hypotheses make predictions about learning outcomes and process-level measures that can be tested empirically. Let us consider predictions for two sequences of instruction: a condition that receives connectional sense-making support before connectional fluency-building support (understanding-first condition) versus a condition that receives connectional fluency-building support before connectional sense-making support (fluency-first condition).

Predictions of learning outcomes

First, let us consider connectional fluency as one of the possible learning outcomes; predictions with respect to connectional understanding are discussed below. The understanding-first hypothesis predicts that students in the understanding-first condition will outperform students in the fluency-first condition on measures of connectional fluency because having acquired connectional understanding equips students with the knowledge that is necessary to benefit from connectional fluency-building support, for instance by helping students to direct their attention to the conceptually relevant aspects of graphical representations. By contrast, the fluency-first hypothesis does not make specific predictions for measures of connectional fluency; students' benefit from connectional fluency-building support does not depend on having previously received connectional sense-making support. Thus, I am investigating the following two competing hypotheses:

Understanding-first connection-fluency H1: The understanding-first condition will outperform the fluency-first condition on measures of connectional fluency.

Fluency-first connection-fluency H0: The fluency-first condition will perform equally well as the understanding-first condition on measures of connectional fluency.

Second, let us consider connectional understanding as another possible learning outcome. The fluency-first hypothesis predicts that students in the fluency-first condition outperform students in the understanding-first condition on measures of connectional understanding because connectional fluency frees cognitive capacities that students can invest in connectional sense-making processes. By contrast, the understanding-first hypothesis does not make specific predictions for measures of connectional understanding; students' benefit from connectional sense-making support is not expected to depend on having previously received connectional fluency-building support. Hence, I am contrasting the following two competing hypotheses:

Fluency-first connection-understanding H1: The fluency-first condition will outperform the understanding-first condition on measures of connectional understanding.

Understanding-first connection-understanding H0: The understanding-first condition will perform equally well as the fluency-first condition on measures of connectional understanding.

Third, let us consider transfer of fractions knowledge as a final measure of learning outcomes. Both hypotheses predict that the optimal sequence of connectional sense-making support and connectional fluency-building support will promote students' learning of robust fractions knowledge that can transfer to novel tasks. According to the understanding-first hypothesis, the understanding-first condition should outperform students in the fluency-first condition on transfer of fractions knowledge. Conversely, the fluency-first condition predicts that the fluency-first condition will outperform students in the understanding-first condition on transfer of domain knowledge. An alternative hypothesis might state that the sequence of connectional sense-making support and connectional fluency-building support does not matter, as long as both are provided. Note that this "null"-hypothesis is consistent with the results from Experiment 4, but inconsistent with both the understanding-first hypothesis and the fluency-first hypothesis. Taken together, I investigate the following contrasting hypotheses:

Understanding-first transfer H1: The understanding-first condition will outperform the fluency-first condition on transfer of fractions knowledge.

Fluency-first transfer H2: The fluency-first condition will outperform the understanding-first condition on transfer of fractions knowledge.

Combination transfer H0: The understanding-first condition and the fluency-first condition will perform equally well on transfer of fractions knowledge.

Predictions of problem-solving behaviors during learning

Both hypotheses make predictions regarding measures of problem-solving behaviors during the learning process. In particular, the number of errors students make while working on connectional sense-making and fluency-building problems as well as the time they spend on these problems are of interest. The understanding-first hypothesis predicts that students in the understanding-first condition will make fewer errors on connectional fluency-building problems than the fluency-first condition because their connectional understanding helps them solve fluency-building problems, which in turn enhances their benefit from fluency-building support.

Likewise, the understanding-first hypothesis predicts that students in the understanding-first condition will spend less time and make fewer errors on fluency-building problems because their connectional understanding enables them to more efficiently direct their attention to relevant aspects of the graphical representations, thereby allowing them to solve fluency-building problems faster. By contrast, the fluency-first condition does not make specific predictions as to the number of errors or the time spent on fluency-building problems. Thus, I contrast the following hypotheses:

Understanding-first errors-fluency H1: The understanding-first condition will make fewer errors on fluency-building problems than the fluency-first condition.

Fluency-first errors-fluency H0: The fluency-first condition and the understanding-first conditions will make the same number of errors on fluency-building problems.

Understanding-first duration-fluency H1: The understanding-first condition will spend less time on fluency-building problems than the fluency-first condition.

Fluency-first duration-fluency H0: The fluency-first condition and the understanding-first conditions will spend the same amount of time on fluency-building problems.

By contrast, the fluency-first condition predicts that students in the fluency-first condition will make fewer errors and will spend less time on sense-making problems than students in the understanding-first condition because connectional fluency frees the cognitive load students need to successfully engage in connectional sense-making processes. The understanding-first condition, on the other hand, does not predict differences between conditions on connectional sense-making problems. Thus, I will compare the following competing hypotheses:

Fluency-first errors-sense H1: The fluency-first condition will make fewer errors on sense-making problems than the understanding-first condition.

Understanding-first errors-sense H0: The understanding-first condition and the fluency-first conditions will make the same number of errors on sense-making problems.

Fluency-first duration-sense H1: The fluency-first condition will spend less time on sense-making problems than the understanding-first condition.

Understanding-first duration-sense H0: The understanding-first condition and the fluency-first conditions will spend the same amount of time on sense-making problems.

Predictions of connection making and conceptual processing while reflecting on learning

Both hypotheses make predictions regarding connection making between graphical representations and conceptual reasoning about fractions concept during the learning process. Specifically, the understanding-first hypothesis predicts that, being equipped with connectional understanding, students in the understanding-first condition will make more connections between graphical representations on connectional fluency-building problems than the fluency-first condition, as assessed by interview data after the learning phase. By contrast, the fluency-first condition does not predict differences between conditions on connection making while students reflect on their work on connectional fluency-building problems. Thus, I contrast the following two hypotheses:

Understanding-first connections-fluency H1: The understanding-first condition will make more connections while reflecting on their work on fluency-building problems than the fluency-first condition.

Fluency-first connections-fluency H0: The fluency-first condition and the understanding-first condition will make the same number of connections while reflecting on their work on fluency-building problems.

Furthermore, the fluency-first hypothesis predicts that, having connectional fluency, students in the fluency-first condition will make more connections between graphical representations than students in the understanding-first condition while reflecting on their work on connectional sense-making problems. However, the understanding-first condition does not predict differences between conditions on connection making while students reflect on their work on connectional sense-making problems. I will therefore compare the following contrasting hypotheses:

Fluency-first connections-sense H1: The fluency-first condition will make more connections while reflecting on their work on sense-making problems than the understanding-first condition.

Understanding-first connections-sense H0: The understanding-first condition and the fluency-first condition will make the same number of connections while reflecting on their work on connectional sense-making problems.

In addition, both the understanding-first and the fluency-first hypotheses imply that the optimal sequence of connectional sense-making and fluency-building support will promote students' conceptual reasoning about fractions. According to the understanding-first hypothesis, one would expect that the understanding-first condition will engage in more conceptual reasoning than the fluency-first condition, as assessed by interview data after the learning phase. Conversely, the fluency-first condition predicts that the fluency-first condition will engage in more conceptual reasoning than the understanding-first condition. An alternative hypothesis might state that the sequence of connectional sense-making support and fluency-building support does not affect conceptual reasoning: as long as both types of support are present, students will engage in conceptual reasoning. Taken together, I will investigate the following hypotheses:

Understanding-first concepts H1: The understanding-first condition will engage in more conceptual reasoning about fractions while reflecting on their work than the fluency-first condition.

Fluency-first concepts H2: The fluency-first condition will engage in more conceptual reasoning about fractions while reflecting on their work than the understanding-first condition.

Combination concepts H0: The understanding-first condition and the fluency-first condition will engage in the same amount of conceptual reasoning about fractions while reflecting on their work.

Predictions of relations between process-level and learning outcome measures

Finally, both the understanding-first and the fluency-first hypotheses make predictions about the relation between these different dependent measures. Differences between conditions on the connectional fluency and connectional understanding should partially explain the differences between conditions on the transfer posttest, while controlling for pretest performance. Similarly, differences in problem-solving behavior are expected to partially explain the differences between conditions on the transfer posttest, while controlling for pretest performance.

4.5.2 Method

To investigate these hypotheses, I conducted a lab-based experiment that contrasted different sequences of connectional sense-making and fluency-building support.

4.5.2.1. Experimental design and procedure

Activity Type	**Understanding-first condition**	**Fluency-first condition**
Test	Pretest: near / far transfer	Pretest: near / far transfer
Tutor: equivalence	***Sense-making support:*** 4 tutor problems	***Fluency-building support:*** 4 tutor problems
Quiz 1: equivalence	Reproduction-understanding, reproduction-fluency	Reproduction-understanding, reproduction-fluency
Tutor: equivalence	***Fluency-building support:*** 4 tutor problems	***Sense-making support:*** 4 tutor problems
Quiz 2: equivalence	Reproduction-understanding, reproduction-fluency	Reproduction-understanding, reproduction-fluency
Tutor: comparison	***Sense-making support:*** 4 tutor problems	***Fluency-building support:*** 4 tutor problems
Quiz 1: comparison	Reproduction-understanding, reproduction-fluency	Reproduction-understanding, reproduction-fluency
Tutor: comparison	***Fluency-building support:*** 4 tutor problems	***Sense-making support:*** 4 tutor problems
Quiz 2: comparison	Reproduction-understanding, reproduction-fluency	Reproduction-understanding, reproduction-fluency

Test	Posttest: near / far transfer	Posttest: near / far transfer
Interview	Retrospective interview on tutor problems	Retrospective interview on tutor problems

Table 23. Experiment 5: Sequence of activities by experimental condition.

Students were randomly assigned to different sequences of connectional sense-making support and connectional fluency-building support. In other words, all students worked on the same tutor problems, but in different orders.

Table 23 details the sequence of assessment problems and tutor problems for each experimental condition. Students first completed a pretest. They then worked on equivalence and comparison topics of the Fractions Tutor. After every four tutor problems, students completed two quiz items (i.e., reproduction-understanding and reproduction-fluency for the given topic). After completing all tutor problems as well as the last set of quiz items, students were given an immediate posttest.

Students in the *understanding-first condition* received connectional sense-making support before connectional fluency-building support. This procedure was implemented for each topic (i.e., equivalence and comparison). Specifically, students in the understanding-first condition first worked on four connectional sense-making problems for equivalent fractions. Next, they worked on four connectional fluency-building problems for equivalent fractions. They then worked on four connectional sense-making problems for fraction comparison, followed by four connectional fluency-building problems for fraction comparison.

By contrast, students in the *fluency-first condition* received connectional fluency-building support before connectional sense-making support, again for each topic. Specifically, students in the fluency-first condition first worked on four connectional fluency-building problems for equivalent fractions, then on four connectional sense-making problems for equivalent fractions. Next, they worked on four connectional fluency-building problems for fraction comparison, followed by four connectional sense-making problems for fraction comparison.

The experiment was conducted in two phases. Due to delayed arrival of the eye-tracking equipment, the first 38 of 74 students participated in the experiment without eye tracking. The remaining 36 students participated in the experiment with eye tracking. The procedure for both phases of the experiment was exactly identical except for the collection of interview data, as detailed below.

Students in phase 2 worked with the SMI RED 250 remote eye-tracking system, which uses an infra-red camera located at the bottom of a regular computer monitor to track students' eye movements (Duchowski, 2007). Therefore, students in phase 2 of the experiment worked on the Fractions Tutor in the same was as students in phase 1 (after following a calibration procedure that took about 1-2 minutes to complete).

4.5.2.3. Participants

Seventy-four students from grades 3-5 participated in the experiment. Students were recruited through advertisements in local newspapers, online bulletin boards, and through flyers distributed in local schools. Sessions were conducted individually in the lab. Students were randomly assigned to the understanding-first condition or to the fluency-first condition.

4.5.2.4. Fractions Tutor versions

Students worked with a subset of the tutor problems used in Experiment 4 (see section 3.4.2). *Connectional sense-making support* makes use of the worked-example principle (Renkl, 2005). Students were first presented with a worked example that uses one of the area models (i.e., circle or rectangle) to demonstrate how to solve a fractions problem. Students completed the last step of the worked-example problem and (while the worked example was still on the screen) were then presented with an equivalent problem in which they have to use the number line. Students had to complete these analogous (i.e., they were open problems, not worked examples). At the end of the problem, students were prompted to relate the two graphical representations to one another. Fig. 13 (see section 3.4.2) shows an example of a connectional sense-making support problem for equivalent fractions, Fig. 14 (see section 3.4.2) shows an example of a connectional sense-making support problem for fraction comparison.

Connectional fluency-building support problems are similar to Kellman et al.'s fluency training for perceptual expertise in connection making (Kellman et al., 2008). Students were presented with a variety of graphical representations and have to sort them into sets of equivalent fractions (see Fig. 15, see section 3.4.3), or order them from smallest to largest, using drag-and-drop (see Fig. 16, section 3.4.3). Students were encouraged to solve the problems by visually estimating the relative size of the fractions, rather than by counting or computationally solving the problems.

4.5.2.5. Test instruments

I assessed learning outcome measures, measures of problem-solving behaviors collected while students worked with the Fractions Tutor, and I conducted retrospective interviews, using the eye-gaze recordings as cues for the interviews for students who participated in the second phase of the experiment.

Learning outcome measures

I assessed *reproduction of connection-making knowledge* based on quiz items with circles, rectangles, and number lines, presented in a format identical to the problems in the Fractions Tutor. Examples of the quiz items can be found in Appendix 8. Specifically, connection-understanding items assessed students' conceptual understanding of connections between graphical representations with regard to equivalent fractions and fraction comparison. For instance, students were asked to choose the correct explanation for why two number lines show equivalent fractions. Connection-fluency items assessed students' fluency in making connections with regard to equivalent fractions and fraction comparison. For example, students were asked to quickly find different graphical representations that show equivalent fractions. The order in which students received the quiz items was counterbalanced. For both measures, I computed accuracy and efficiency scores. Accuracy was computed as the proportion of correct responses to the maximum number correct responses. To assess students' efficiency in solving quiz items, I took into account the speed with which students solved the quiz items, following (van Gog & Paas, 2008):

$$\text{efficiency} = \frac{\mathrm{Z}\,(\text{proportion correct}) - \mathrm{Z}\,(\text{time on quiz items})}{\sqrt{2}} \tag{7}$$

Positive efficiency scores indicate higher efficiency at solving quiz items correctly, and negative efficiency scores indicate lower efficiency at solving quiz items correctly, compared to the relative mean of the sample. An efficiency score of 0 indicates average efficiency with respect to the sample.

I assessed students' *transfer of fractions knowledge* based on equivalent pretests and posttests. This measure corresponds to transfer items from the conceptual knowledge test used in Experiment 4, while reflecting the smaller selection of topics covered in Experiment 5. The transfer test included items on equivalence and comparison without graphical representations. Example test items are shown in Appendix 8. I computed accuracy and efficiency scores for the test.

Tutor log data

The Fractions Tutor logs all of the students' interactions during problem solving. The analysis of the tutor log data is based on a knowledge component model, summarized in Table 24. The knowledge component model was the result of a comparison of multiple alternative knowledge component models. Fig. 29 and Fig. 30 provide examples for the knowledge components for the equivalent fractions sense-making problems and fraction comparison sense-making problems, respectively.

To investigate the predictions made by the understanding-first and fluency-first hypotheses, I compared conditions on several measures based on this knowledge component model. First, I investigated trends in students' *first-attempt errors* and *step durations*. To do so, I used "learning curves" provided by the DataShop web service (Koedinger et al., 2010), which depict the average error rate or step duration (across students and knowledge components) as a function of the amount of practice (i.e., the number of opportunities a student has to apply a given knowledge component). Following standard practice in Cognitive Tutors research (Koedinger et al., 2010, 2012), I considered each step in a given tutor problem as a learning opportunity for the particular knowledge component involved in the step. I considered a step in the problem to be correct if the student solved it without hints and without errors (i.e., if the student's first action on the step was a correct attempt at solving, as opposed to an error or a hint request). I consider the total time spent on a step involving a given knowledge component as the duration of the step. I expect that, if learning occurs, the number of errors students make on consecutive learning opportunity decrease, and that they spend a decreasing amount of time on consecutive learning opportunity. In other words, decreasing learning curves indicate learning.

Second, I investigated whether the number of *first-attempt errors* and *total step duration* mediate the effects of condition on students' accuracy and efficiency on the quiz items and transfer tests. I computed the number of first-attempt errors as the sum of first actions on a given step was that were incorrect attempts across (1) equivalence knowledge components in fluency-building problems (i.e., knowledge components 1-3 in Table 24), (2) equivalence knowledge components in sense-making problems (i.e., knowledge components 4-10 in Table 24), (3) comparison knowledge components in fluency-building problems (i.e., knowledge components 11-13 in Table 24), (4) comparison knowledge components in sense-making problems (i.e., knowledge components 14-21 in Table 24). Likewise, I computed the step duration as the total amount of time students spent on a given step across the same knowledge component types listed in Table 24.

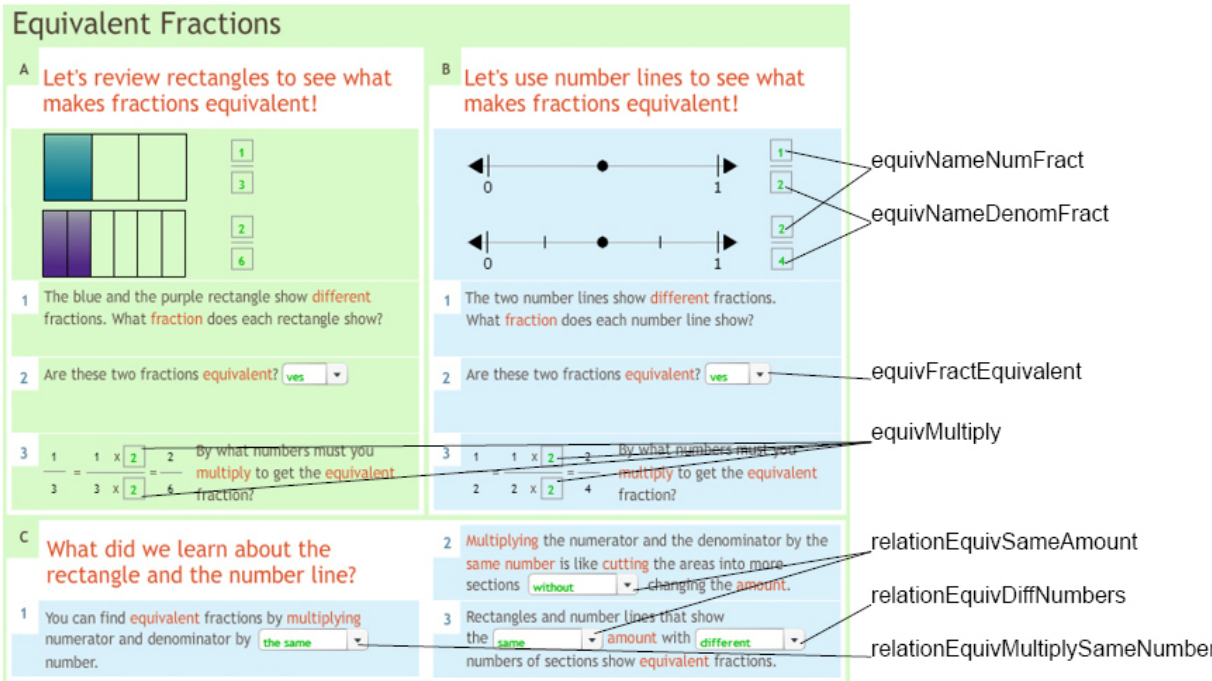


Fig. 29. Experiment 5: Example knowledge components for equivalent fraction sense-making problems.

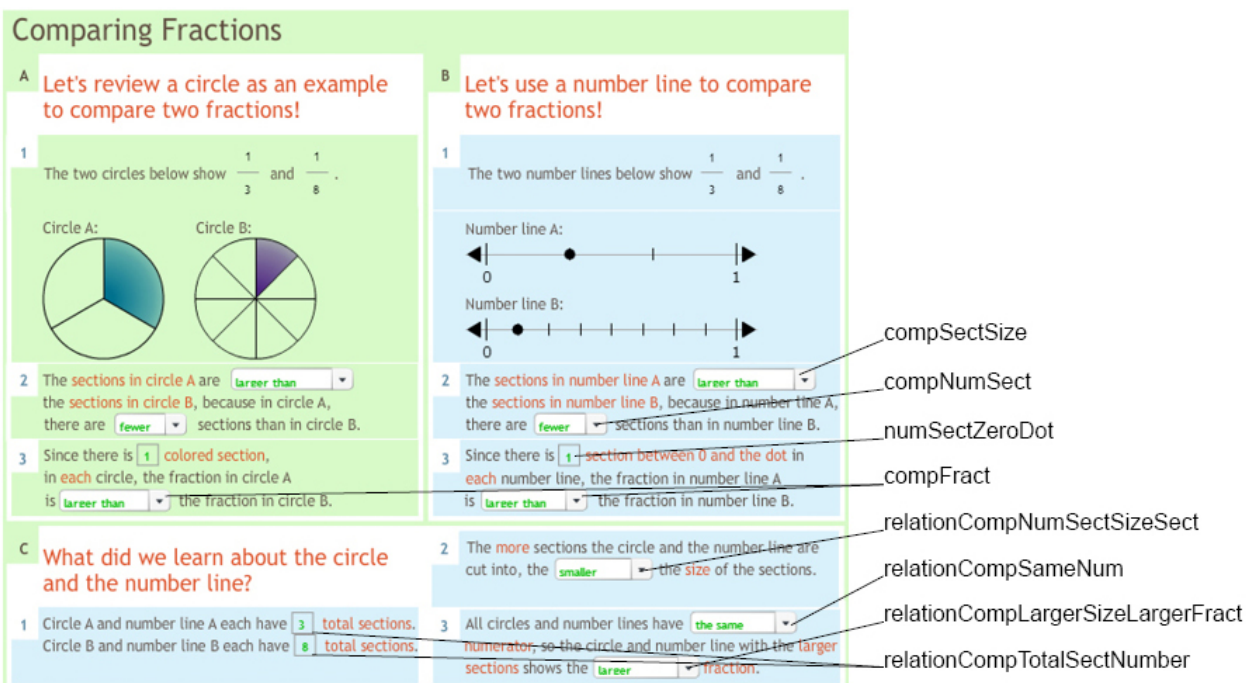


Fig. 30. Experiment 5: Example knowledge components for fraction comparison sense-making problems.

Knowledge component	Knowledge component type	Description
1. equivDrag_circle	Equivalence fluency-building	Dragging-and-dropping equivalent circle (fluency-building problem)
2. equivDrag_rect	Equivalence fluency-building	Dragging-and-dropping equivalent rectangle (fluency-building problem)
3. equivDrag_NL	Equivalence fluency-building	Dragging-and-dropping equivalent number line (fluency-building problem)
4. equivFractEquivalent	Equivalence sense-making	Judging whether two fraction representations are equivalent (sense-making problem)
5. equivMultiply	Equivalence sense-making	Multiplying the numerator or denominator of a given symbolic fraction to find an equivalent fraction (sense-making problem)
6. equivNameDenomFract	Equivalence sense-making	Entering the denominator of an equivalent fraction (sense-making problem)
7. equivNameNumFract	Equivalence sense-making	Entering the numerator of an equivalent fraction (sense-making problem)
8. relationEquivDiffNumbers	Equivalence sense-making	Reasoning about equivalent fractions showing the same amount with different numbers (sense-making problem)
9. relationEquivMultiplySameNumber	Equivalence sense-making	Reasoning about multiplying numerator and denominator by the same number to find equivalent fractions (sense-making problem)
10. relationEquivSameAmount	Equivalence sense-making	Reasoning about equivalent fractions showing the same amount (sense-making problem)
11. compDrag_circle	Comparison fluency-building	Dragging-and-dropping circle that is smaller or larger than another (fluency-building problem)
12. compDrag_rect	Comparison fluency-building	Dragging-and-dropping rectangle that is smaller or larger than another (fluency-building problem)
13. compDrag_NL	Comparison fluency-building	Dragging-and-dropping number line that is smaller or larger than another (fluency-building problem)
14. compFract	Comparison sense-making	Indicating which fraction representation is larger or smaller than another (sense-making problem)
15. compNumSect	Comparison sense-making	Comparing the number of sections two fraction representations are partitioned into (sense-making problem)
16. compSectSize	Comparison sense-making	Comparing the relative size of sections in two fraction representations (sense-making problem)
17. numSectZeroDot	Comparison sense-making	Entering the fraction shown on the number

		line (sense-making problem)
18. relationCompLargerSizeLargerFract	Comparison sense-making	Reasoning about the fraction representation with the larger sections showing the larger fraction, given that the numerators are the same (sense-making problem)
19. relationCompNumSectSizeSect	Comparison sense-making	Reasoning about the number of sections a fraction representation is partitioned into being inversely related to the size of the sections (sense-making problem)
20. relationCompSameNum	Comparison sense-making	Reasoning about the number of sections of two fraction representations being the same if the numerators are the same (sense-making problem)
21. relationCompTotalSectNumber	Comparison sense-making	Reasoning about the number of total sections two fraction representations are partitioned into (sense-making problem)

Table 24. Experiment 5: Knowledge component model.

Retrospective interviews

In phase 1 of the experiment (i.e., in the first phase without eye tracking), students were interviewed about their problem-solving procedure on four randomly selected tutor problems. Specifically, one of each problem type (i.e., equivalence sense-making support, equivalence fluency-building support, comparison sense-making support, comparison fluency-building support, see overview in Table 24) was randomly selected for the retrospective interview. On these randomly selected problems, the interviewer asked previously specified questions about how the student solved each step in the problem, based on detailed notes about students' interactions. For instance, the interviewer might ask: “In this step (points at step B2 in Fig. 29), you immediately selected ‘yes’ from the menu, that the two fractions in the number lines are equivalent. How did you solve that step?”

In phase 2 of the experiment (i.e., the second phase with eye-gaze recordings as cues), four tutor problems were randomly selected in the same way as in phase 1. For each of these randomly selected problems, the interviewer played back the recorded eye-gaze behaviors for each selected problem to the student. The eye-gaze recordings depict the student's eye-gaze focus as a circle, overlaid with a background-screen recording showing the Fractions Tutor problem and the student's interactions with the problem. In replaying the eye-gaze recording to the student, the interviewer first explained to the student what the eye-gaze circle meant, and then paused after each completed step for an interview question. Then, the interviewer asked about the student's problem-solving behavior in the same was as in phase 1, but skipping the description of the student's interaction (i.e., in the earlier example, the interviewer would *not* say "In this step [points at step B2 in Fig. 29], you immediately selected 'yes' from the menu, that the two fractions in the number lines are equivalent," since that information was redundant with the eye-gaze recording). Instead, the interviewer might for example ask, after replaying the eye-gaze recording for step B2 in Fig. 29, "how did you solve this step?"

Code	Description
Only for sense-making problems: 1. Representation connections	If the student is talking about two different graphical representations (e.g., circle – rectangle, circle – number line, rectangle – number line), not if the student is talking about the same graphical representations (e.g., circle – circle, rectangle – rectangle, number line – number line)
1.1. Representation-surface	If the student makes a surface-level connection between two different graphical representations (e.g., based on color, shape, or some other feature that is not conceptually relevant). *Example:* "Like when you show it (points at rectangle B) this rectangle and the number lines, the number lines have the lines, and that's for the rectangle." (from an equivalence sense-making problem)
1.2. Representation-concept-incorrect	If the student refers to a structural feature (e.g., number of total sections, number of shaded sections) of two different graphical representations (i.e., circle – rectangle, circle – number line) but does so incompletely or incorrectly. *Example:* "I saw that the they were both the same number, they're both the same." (from an equivalence sense-making problem)
1.3. Representation-concept-correct	If the student refers to a structural feature correctly (e.g., number of total sections, number of shaded sections) of two different graphical representations (i.e., circle – rectangle, circle – number line). *Example:* "Three, em, the circles and the number lines show the same thing, cause this (points at number line B) is like one third, but it's still nine, so three is one third of nine. And then this one (points at second question in C3), em, it was just em, that they have like different numbers of em, sections, so I saw numbers of sections and I just like saw it, they're different sections, so numbers of sections." (from an equivalence sense-making problem)
Only for fluency-building problems: 2. Representation-fluency	If the student is talking about two different graphical representations (e.g., circle – rectangle, circle – number line, rectangle – number line), not if the student is talking about the same graphical representations (e.g., circle – circle, rectangle – rectangle, number line – number line). The student needs to refer to either an intuitive sense of the two representations looking alike, by visually estimating, or needs to correctly explain the connection based on concepts (e.g., numerator and denominator, as in 1.3). *Example:* "Em, well I was sort of looking at em, at how much there was, like I noticed that this (points at circle 1) equals this (points at given circle for slot 1), so this (points at circle 1) had to be with this one (points at given circle for slot 1) because they're the same amount." (from an equivalence fluency-building problem)
For all problems: 3. Concept-correct	If the student explains a fractions concept correctly, without relating to graphical representations, while relating to one graphical representation (e.g., only a circle), or while relating to the same graphical representations (e.g., circle – circle) *Example:* "Em, well you don't really change, I just knew that you don't really change the amount here (points at circles), because it looks like this one (points at

	circle A) is the same shape as that one (points at circle B), so they're basically like the same amount." (from an equivalence sense-making problem)

Table 25. Experiment 5: Coding scheme for retrospective interviews.

The protocols obtained from the retrospective interviews were coded for conceptual processing and surface-level processing of connections between multiple graphical representations, as well as for conceptual understanding of fractions (regardless of connections between multiple graphical representations). Table 25 details the criteria for each of these codes and example utterances for each code. Fig. 31 shows the decision tree used by the coders to facilitate the coding procedure. I applied the same coding scheme to the retrospective interviews obtained in phases 1 and 2. However, due to the different sources of information available to the students as they responded to the interview question (i.e., a verbal description of the interaction versus eye-gaze recordings), I analyze the results from phases 1 and 2 separately.

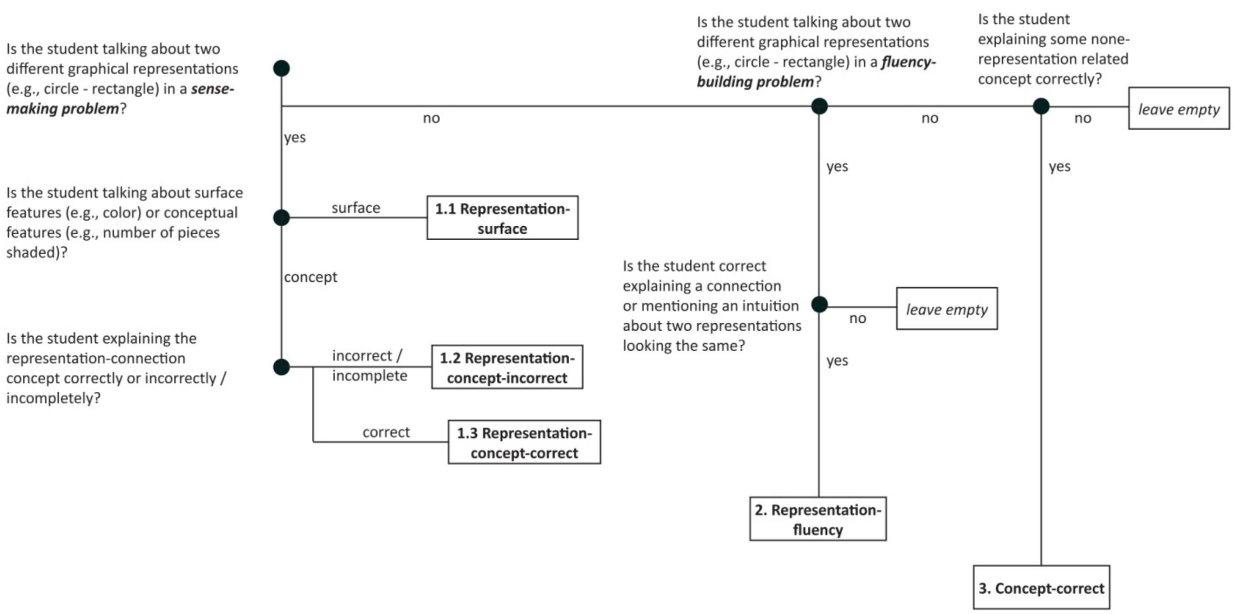


Fig. 31. Experiment 5: Decision tree to apply the coding scheme for retrospective interviews.

4.5.3 Results

Five students were excluded from the analysis. One student was excluded because he did not complete both topics of the Fractions Tutor. Four students were excluded because they were statistical outliers at the pretest. Table 26 shows the number of included and excluded students per condition.

	included	excluded

understanding-fist	37	3
fluency-first	32	2

Table 26. Experiment 5: Number of students included and excluded by condition.

Taken together, the data from $N = 69$ students were analyzed ($n = 37$ in the understanding-first condition, $n = 32$ in the fluency-first condition). I report partial eta-squared as measures of effect size, with an effect size of η^2 of .01 corresponding to a small effect, .06 to a medium effect, and .14 to a large effect (Cohen, 1988). Table 27 gives an overview of the results on the learning outcome measures. Table 28 provides an overview of the results on the process-level measures.

measure	test time	effect	significant	F/t-value	adj. *p*-value	effect size
Quiz: connection-fluency accuracy		condition	marginally	$F(1,65) = 3.34$	$p < .10$	$\eta^2 = .05$
		quiz time	no	$F < 1$		
		condition * quiz time	no	$F(1,65) = 1.42$	$p > .10$	
	quiz 1	understanding-first ~ fluency-first	no	$F < 1$		
	quiz 2	understanding-first > fluency-first	yes	$F(1,65) = 4.52$	$p < .05$	$\eta^2 = .07$
Quiz: connection-fluency efficiency		condition	yes	$F(1,65) = 12.14$	$p < .01$	$\eta^2 = .16$
		quiz time	no	$F(1,65) = 1.17$	$p > .10$	
		condition * quiz time	yes	$F(1,65) = 6.55$	$p < .05$	$\eta^2 = .09$
	quiz 1	fluency-first > understanding-first	yes	$F(1,65) = 11.34$	$p < .01$	$\eta^2 = .15$
	quiz 2	fluency-first > understanding-first	marginally	$F(1,65) = 2.82$	$p < .10$	$\eta^2 = .04$
Quiz: connection-understanding accuracy		condition	no	$F < 1$		
		quiz time	no	$F < 1$		
		condition * quiz time	no	$F < 1$		
Quiz: connection-understanding efficiency		condition	no	$F < 1$		
		quiz time	no	$F < 1$		
		condition * quiz time	no	$F < 1$		
Test: transfer accuracy		condition	no	$F < 1$		
		test time	yes	$F(1,66) = 3.76$	$p < .01$	$\eta^2 = .12$
		condition * test time	no	$F < 1$		
	posttest	understanding-first > fluency-first	marginally	$F(1,65) = 3.05$	$p < .10$	$\eta^2 = .05$
Test: transfer efficiency		condition	no	$F < 1$		
		test time	yes	$F(1,66) = 8.66$	$p < .01$	$\eta^2 = .12$
		condition * test time	no	$F < 1$		

Table 27. Experiment 5: Results on learning outcome measures.

measure	problem type	effect	significant	F/χ²-value	adj. *p*-value	effect size
error rate	equivalence-sense	understanding-first < fluency-first	no	$F(1, 66) = 2.61$	$p > .10$	
	comparison-sense	understanding-first ~ fluency-first	marginally	$F(1, 66) = 2.80$	$p < .10$	$\eta^2 = .04$
	equivalence-fluency	understanding-first < fluency-first	marginally	$F(1, 66) = 2.79$	$p < .10$	$\eta^2 = .04$
	comparison-fluency	understanding-first ~ fluency-first	no	$F(1, 66) = 2.04$	$p > .10$	
step duration	equivalence-sense	understanding-first ~ fluency-first	no	$F < 1$		
	comparison-sense	understanding-first ~ fluency-first	no	$F < 1$		
	equivalence-fluency	understanding-first ~ fluency-first	no	$F < 1$		
	comparison-fluency	understanding-first ~ fluency-first	no	$F < 1$		
representation-connections (phase 1 interviews)	sense-making problems	---	---			
	fluency-building problems	understanding-first ~ fluency-first	no	$\chi^2 < 1$		
	all tutor problems	fluency-building problems > sense-making problems	yes	$\chi^2(1, N = 187) = 88.99$	$p < .01$	
conceptual reasoning (phase 1 interviews)	all tutor problems	understanding-first > fluency-first	yes	$\chi^2(1, N = 351) = 10.60$	$p < .01$	

Table 28. Experiment 5: Results on process-level measures.

4.5.3.1. Effects on quiz: Connection-fluency and connection-understanding

		Understanding-first	Fluency-first
Accuracy	Quiz 1: connection-understanding	.49 (.36)	.39 (.38)
	Quiz 2: connection-understanding	.41 (.33)	.45 (.41)
	Quiz 1: connection-fluency	.62 (.35)	.60 (.32)
	Quiz 2: connection-fluency	.65 (.34)	.50 (.31)
Efficiency	Quiz 1: connection-understanding	-.04 (12.99)	.12 (10.05)
	Quiz 2: connection-understanding	.41 (.33)	.45 (.41)
	Quiz 1: connection-fluency	-6.32 (21.32)	6.61 (8.61)
	Quiz 2: connection-fluency	-1.31 (9.47)	1.71 (5.62)

Table 29. Experiment 5: Means and standard deviations (in parentheses) by condition on connection-understanding and connection-fluency measures per quiz time.

Table 29 shows the means and standard deviations for the understanding-first and fluency-first conditions on the quiz measures: accuracy and efficiency on the reproduction-understanding and the reproduction-fluency scales of the quizzes at quiz times 1 and 2. The scores were averaged across the equivalence and comparison topics (see Table 19 for an overview of the procedure), because the measure of interest regards fluency and understanding, but not knowledge about equivalence and comparison.

To investigate the *understanding-first connection-fluency hypothesis H1* (that the understanding-first condition will outperform the fluency-first condition on measures of fluency in making connections), I conducted a repeated measures ANCOVA with accuracy on the pretest and time spent on the Fractions Tutor as covariates, and *accuracy on reproduction-fluency* at quiz times 1 and 2 as repeated, dependent measures. Fig. 32 shows students' accuracy on the reproduction-fluency quizzes at quiz times 1 and 2. There was a marginally significant main effect of condition, $F(1,65) = 3.34$, $p < .10$, $\eta^2 = .05$, but no main effect of quiz time ($F < 1$) nor an interaction of quiz time with condition, $F(1,65) = 1.42$, $p > .10$. Posthoc comparisons revealed no significant differences between conditions at quiz time 1 ($F < 1$), but a significant advantage of the understanding-first condition over the fluency-first condition at quiz time 2, $F(1,65) = 4.52$, $p < .05$, $\eta^2 = .07$.

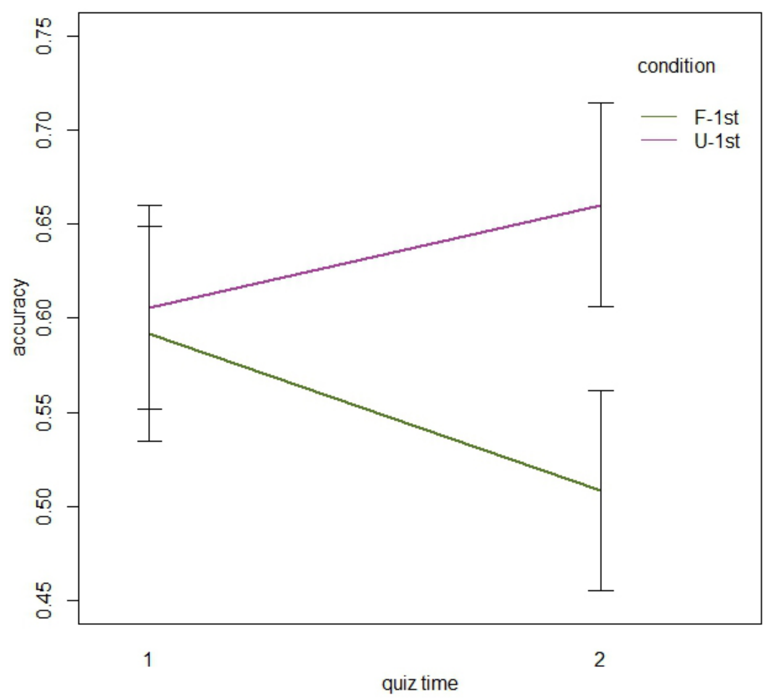


Fig. 32. Experiment 5: Accuracy on reproduction-fluency quizzes by condition (F-1st = fluency-first, U-1st = understanding-first) by quiz time.

A repeated measures ANCOVA with efficiency on the pretest and time spent on the Fractions Tutor as covariates, and *efficiency on reproduction-fluency* at quiz times 1 and 2 as repeated, dependent measures showed a significant main effect of condition, $F(1,65) = 12.14$, $p < .01$, $\eta^2 = .16$, and a significant interaction of quiz time with condition, $F(1,65) = 6.55$, $p < .05$, $\eta^2 = .09$, but no significant effect of quiz time, $F(1,65) = 1.17$, $p > .10$. Fig. 33 shows students' efficiency on the reproduction-fluency quizzes at quiz times 1 and 2. Posthoc comparisons show a significant advantage of the fluency-first condition at quiz time 1, $F(1,65) = 11.34$, $p < .01$, $\eta^2 = .15$, but only a marginally significant advantage of the fluency-first condition at quiz time 2, $F(1,65) = 2.82$, $p < .10$, $\eta^2 = .04$.

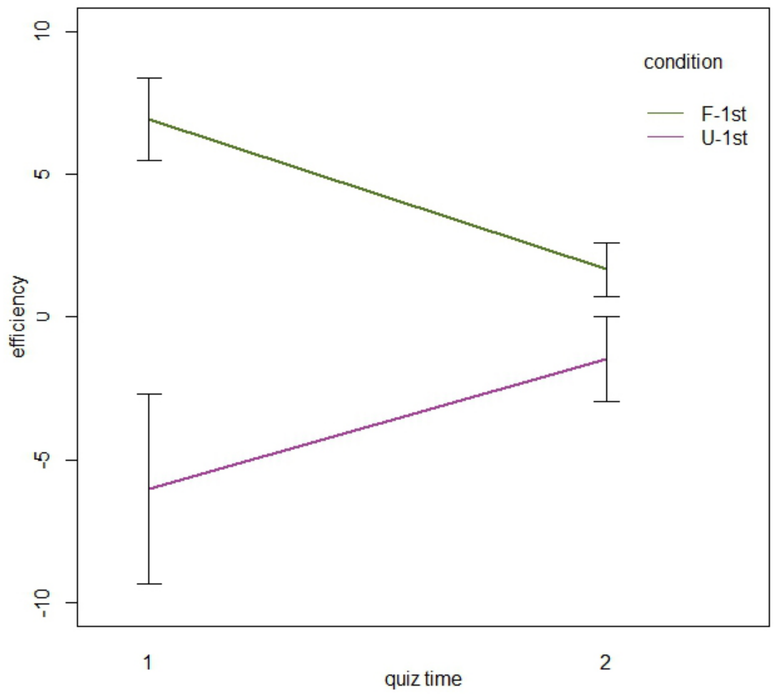


Fig. 33. Experiment 5: Efficiency on reproduction-fluency quizzes by condition (F-1st = fluency-first, U-1st = understanding-first) by quiz time.

To investigate the *fluency-first connection-understanding hypothesis H1* (that the fluency-first condition will outperform the understanding-first condition on measures of conceptual understanding of connections), I conducted a repeated measures ANCOVA with pretest and time spent on the Fractions Tutor as covariates, and *accuracy on reproduction-understanding* at quiz times 1 and 2 as repeated, dependent measures. There was no significant main effect of condition, quiz time, nor an interaction of quiz time with condition (Fs < 1). A repeated measures ANCOVA with efficiency on the pretest and time spent on the Fractions Tutor as covariates, and *efficiency on reproduction-understanding* at quiz times 1 and 2 as repeated, dependent measures showed no significant main effect of condition, quiz time, nor an interaction of quiz time with condition (*F*s < 1).

Taken together, the results provide partial support for the *understanding-first connection-fluency hypothesis H1*. With regard to accuracy on connection-fluency, the results support the understanding-first connection-fluency hypothesis H1: students in the understanding-first condition outperform students in the fluency-first condition significantly at quiz time 2, that is, after having received fluency-building support. However, with regard to efficiency on connection-fluency, the results show the opposite pattern: students in the fluency-first condition significantly outperform students in the understanding-first condition at quiz time 1, although this difference becomes smaller at quiz time 2, that is, after students in the understanding-first condition received fluency-building support, but remains marginally significant. Thus, the *fluency-first connection-fluency hypothesis H0* can only be rejected in favor of the understanding-first connection-fluency hypothesis H1 with regard to accuracy measures, but not with regard to efficiency measures.

Overall, the results do not support the *fluency-first connection-understanding hypothesis H1*. There were no differences between conditions on either accuracy or efficiency of connection-understanding. Thus, the *understanding-first connection-understanding hypothesis H0* cannot be rejected.

4.5.3.2. Effects on posttest: Transfer of knowledge

Table 30 shows the means and standard deviations for the understanding-first and fluency-first conditions on the transfer test measures: accuracy and efficiency at test times 1 and 2.

		Understanding-first	**Fluency-first**
Accuracy	Transfer pretest	.45 (.35)	.53 (.34)
	Transfer posttest	.51 (.36)	.47 (.32)
Efficiency	Transfer pretest	.34 (.44)	.42 (.40)
	Transfer posttest	.51 (.48)	.63 (.44)

Table 30. Experiment 5: Means and standard deviations (in parentheses) by condition on transfer tests per test time.

To investigate the *understanding-first transfer hypothesis H1* (that the understanding-first condition will outperform the fluency-first condition on transfer) and the *fluency-first transfer hypothesis H2* (that the fluency-first condition will outperform the understanding-first condition on transfer), respectively, I computed a repeated measure ANCOVA with time spent on the Fractions Tutor as covariate and test time (pretest and posttest) the repeated factor, and *accuracy on transfer* as dependent measure. Fig. 34 shows students' accuracy on the transfer test at the pretest and the posttest. The results show a marginally significant interaction of test time with condition, $F(1,66) = 3.76$, $p < .10$, $\eta^2 = .05$, but no significant main effects of condition or test (Fs < 1). Posthoc comparisons on the posttest with time spent on the Fractions Tutor and pretest as covariates shows a marginally significant advantage of the understanding-first condition at the posttest, $F(1,65) = 3.05$, $p < .10$, $\eta^2 = .05$.

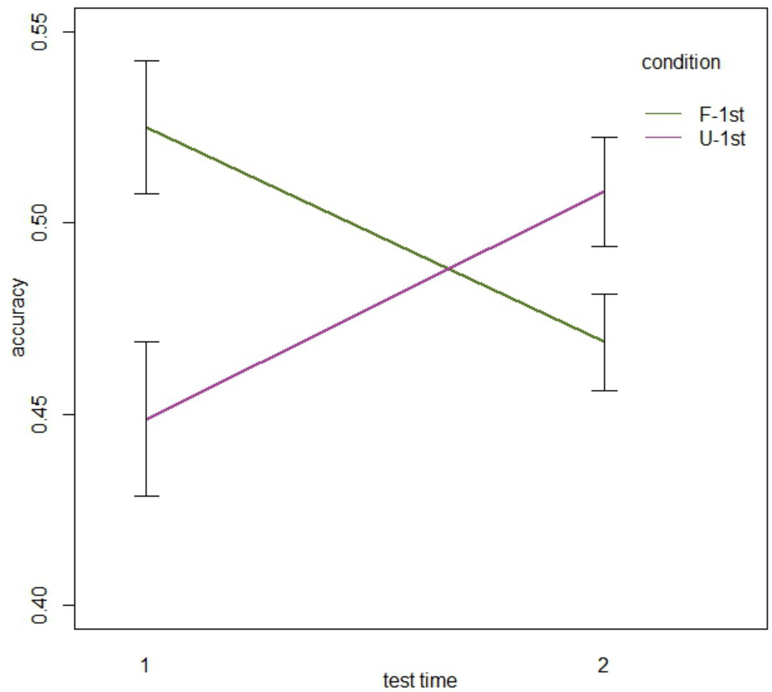


Fig. 34. Experiment 5: Accuracy on transfer by condition (F-1st = fluency-first, U-1st = understanding-first) by test time.

A repeated measures ANCOVA with time spent on the Fractions Tutor as covariate and test time (pretest and posttest) the repeated factor, and *efficiency on transfer* as dependent measure showed a significant main effect of test time, $F(1,66) = 8.66$, $p < .01$, $\eta^2 = .12$, but no significant main effect of condition nor a significant interaction of test time with condition (Fs < 1). Fig. 35 shows students' efficiency on the reproduction-fluency quizzes at quiz times 1 and 2.

Taken together, the results provide partial support for the understanding-first transfer hypothesis H1, but they do not support the fluency-first transfer hypothesis H2. With regard to accuracy on transfer, the understanding-first condition marginally significantly outperforms the fluency-first condition. It is interesting to note that only the understanding-first condition improved from pretest to posttest on accuracy on transfer; the fluency-first condition worsened from pretest to posttest. However, both conditions equally improved on efficiency on transfer. Taken together, the combination transfer hypothesis H0 can only be rejected in favor of the understanding-first transfer hypothesis H1 when considering accuracy on transfer, but not with regard to efficiency.

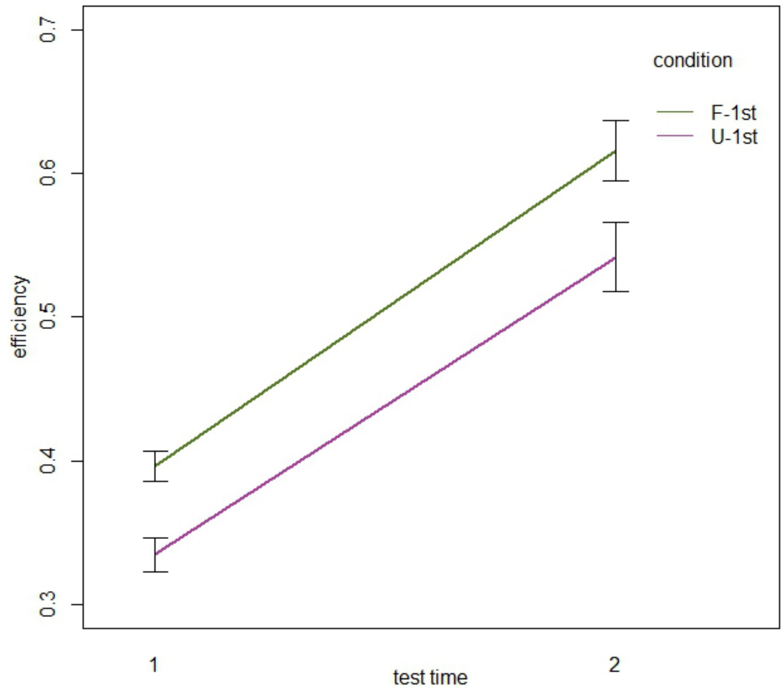


Fig. 35. Experiment 5: Efficiency on transfer by condition (F-1st = fluency-first, U-1st = understanding-first) by test time.

4.5.3.3. Effects on learning curves: Rates of learning

Table 31 shows the means and standard deviations for the understanding-first and fluency-first conditions on the error rates and durations per problem type.

	Problem type	**Understanding-first**	**Fluency-first**
Error rates	equivalence-sense	.17 (.09)	.17 (.08)
	comparison-sense	.28 (.11)	.29 (.11)
	equivalence-fluency	.35 (.14)	.36 (.13)
	comparison-fluency	.17 (.10)	.24 (.10)
Step duration	equivalence-sense	697.06 s (319.97 s)	635.85 s (364.26 s)
	comparison-sense	659.56 s (397.44 s)	573.93 s (219.41 s)
	equivalence-fluency	180.46 s (85.50 s)	180.27 s (98.88 s)
	comparison-fluency	23.57 s (14.09 s)	25.4 s (11.94 s)

Table 31. Experiment 5: Means and standard deviations (in parentheses) by condition on error rates and durations (in seconds) per problem type.

To examine the *understanding-first errors-fluency hypothesis H1* (that the understanding-first condition will make fewer errors on connectional fluency-building support than the fluency-first condition), and the *fluency-first errors-sense hypothesis H1* (that the fluency-first condition will make fewer errors on connectional sense-making support than the understanding-first condition), I computed MANCOVAs with accuracy on the transfer pretest as covariate, and the estimated error rates, as outputted by the DataShop web service, for equivalence sense-making knowledge components, equivalence fluency-building knowledge components, comparison sense-making knowledge components, and comparison fluency-building knowledge components (see Table 24) as dependent measures. There was a marginally significant difference between conditions on comparison sense error rates, $F(1, 66) = 2.80$, $p < .10$, $\eta^2 = .04$, and on equivalence-fluency error rates, $F(1, 66) = 2.79$, $p = .10$, $\eta^2 = .04$, such that the understanding-first condition evidenced lower error rates. There were no significant differences on comparison-fluency error rates, $F(1, 66) = 2.61$, $p > .10$, or on equivalence-sense error rates, $F(1, 66) = 2.04$, $p > .10$.

This finding provides partial support for the *understanding-first errors-fluency hypothesis H1*. Earlier in the learning sequence (that is, on equivalence-fluency problems), the understanding-first condition shows slightly lower error rates than the fluency-first condition. Thus, when considering problem-solving behaviors earlier during the learning sequence, the fluency-first errors-fluency hypothesis H0 can be rejected in favor of the understanding-first errors-fluency hypothesis H1. Whether the fact that this difference was not replicated on the comparison-fluency problems is due to the fact that they occurred later in the learning process (perhaps due to fatigue) or because of attributes of the comparison topic (perhaps due to the fact that these problems appeared to be easier for students, as indicated by lower error rates shown in Table 31) cannot be answered with the current experiment, since equivalence problems were always presented before comparison problems.

It is interesting that this difference does not only occur on fluency-building problems, as the understanding-first hypothesis predicted, but also on sense-making problems. Fig. 36 shows the learning curve for the error rate averaged across comparison sense-making knowledge components. Recall that a decreasing error rate indicates learning. As the standard errors in Fig. 36 indicate, this difference is reliable after the third attempt per knowledge component. These results show that students in the understanding-first condition learn more efficiently than students in the fluency-first condition while working on comparison sense-making problems. This finding stands in contrast to the *fluency-first errors-sense hypothesis H1*. Rather than promoting students' benefit from sense-making problems, receiving fluency-building support first seems to decrease students' benefit from subsequent sense-making support. Again, this difference is only apparent later during the learning sequence (that is, on comparison problems). As with the fluency-building problems, one may reason that an apparent disadvantage of fluency takes a while to manifest itself in an increased error rate during problem solving.

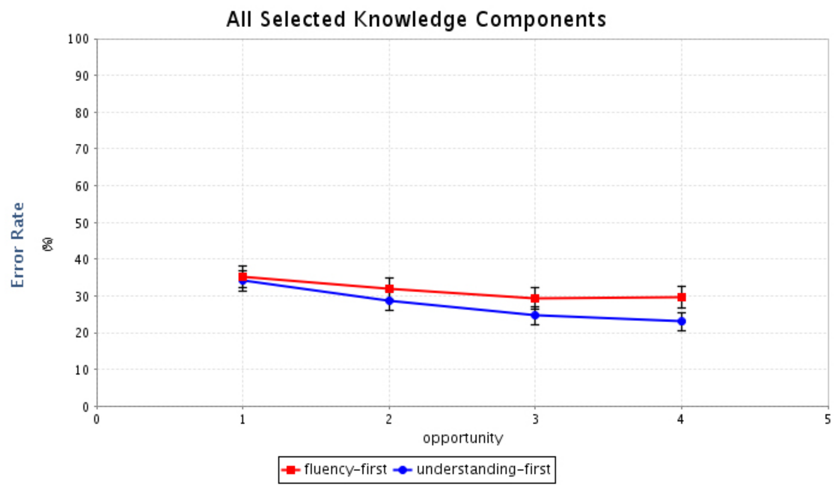


Fig. 36. Experiment 5: Learning curves by condition across comparison sense-making knowledge components. Bars show standard errors.

To investigate the *understanding-first duration-fluency hypothesis H1* (that the understanding-first condition will spend less time on fluency-building problems than the fluency-first condition), and the *fluency-first duration-sense hypothesis H1* (that the fluency-first condition will spend less time on sense-making problems than the understanding-first condition), I computed MANCOVAs with accuracy on the transfer pretest as covariate, and step duration for equivalence sense-making knowledge components, equivalence fluency-building knowledge components, comparison sense-making knowledge components, and comparison fluency-building knowledge components (see Table 24) as dependent measures. There were no significant differences between conditions on either dependent measure (Fs < 1). Thus, neither the fluency-first duration-fluency hypothesis H0 nor the understanding-first duration-sense hypothesis H0 can be rejected.

4.5.3.4. Effects on retrospective interviews: Connection making and conceptual reasoning

At this time, only the interview data collected during phase 1 has been completely transcribed and coded. Table 32 shows the frequencies (i.e., the number of utterances) of representation connections and conceptual utterances obtained during phase 1 of the experiment by the codes described in Table 25. The interrater reliability between two independent coders was substantial with $\kappa = .66$.

	Understanding-first	**Fluency-first**
Representation-surface	1	0
Representation-concept-incorrect	5	1
Representation-concept-correct	12	11
Representation-fluency	76	82
Concept-correct	206	145

Table 32. Experiment 5: Frequencies of utterances coded as representation connections and conceptual reasoning (see Table 25 for the coding scheme).

To test the *understanding-first connections-fluency hypothesis H1* (that the understanding-first condition will make more connections while reflecting on their work on fluency-building problems than the fluency-first condition), I computed a chi-square test on the representation-fluency utterances. The results show no significant difference between conditions ($\chi^2 < 1$). Thus, the fluency-first connections-fluency hypothesis H0 cannot be rejected.

Given that there were almost no representation-concept-correct connections (only 23 in total, see Table 32), a chi-square test to investigate the *fluency-first connections-sense hypothesis H1* (that the fluency-first condition will make more connections while reflecting on their work on sense-making problems than the understanding-first condition) is not warranted. Thus, the understanding-first connections-sense hypothesis H0 cannot be rejected.

To test the *understanding-first concepts hypothesis H1* (that the understanding-first condition will engage in more conceptual reasoning about fractions while reflecting on their work than the fluency-first condition) and the *fluency-first concepts hypothesis H2* (that the fluency-first condition will engage in more conceptual reasoning about fractions while reflecting on their work than the understanding-first condition), I computed a chi-square test on the concept-correct utterances. The results show a significant difference between conditions, $\chi^2(1, N = 351) = 10.60, p < .01$, such that the understanding-first condition engages in significantly more conceptual reasoning about fractions than the fluency-first condition. Thus, the combination-concepts hypothesis H0 and the fluency-first concepts hypothesis H2 can be rejected in favor of the understanding-first concepts hypothesis H1.

It is interesting to note that the fluency-building problems elicited more connection-making utterances than the sense-making problems. A chi-square test comparing the frequencies of all representation-connections to the frequency of all fluency-connections showed that this difference was significantly reliable, $\chi^2(1, N = 187) = 88.99$, $p < .01$.

Taken together, the results from the retrospective interviews obtained in phase 1 of the experiment provide support for the understanding-first hypothesis with regard to conceptual reasoning: students in the understanding-first condition engage in more conceptual reasoning than students in the fluency-first condition. However, there was no difference between conditions with regard to connection making. The sense-making problems elicited little connection making, whereas the fluency-building problems elicited significantly more connection making in both conditions. This latter finding might explain why both types of support are needed. Connectional sense-making support promotes conceptual reasoning about domain-relevant concepts, whereas fluency-building support promotes connection making between the different graphical representations. It will be interesting to find out whether the same results hold true for the retrospective interview data collected in phase 2 of the experiment.

4.5.3.5. Relations between dependent measures

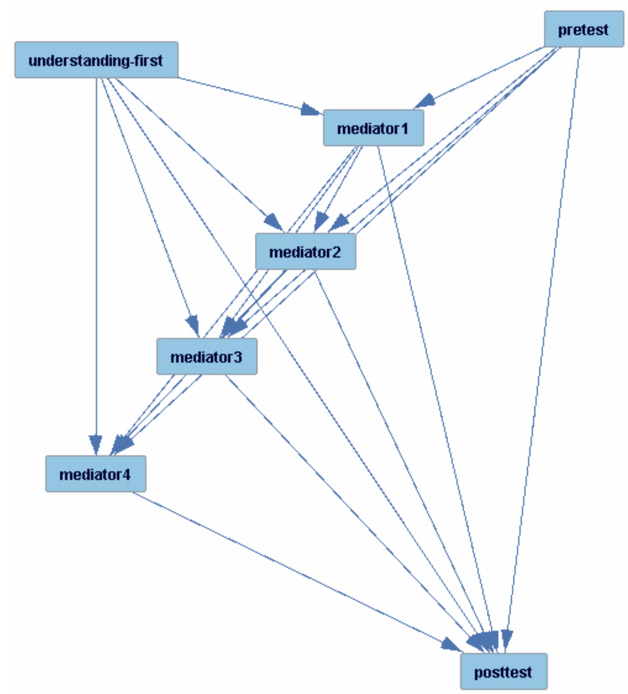


Fig. 37. Experiment 5: Example mediation model compatible with the experimental design.

Next, I investigated whether the differences between conditions on problem-solving behaviors and on the quiz items statistically "explain" (i.e., mediate) the advantage of the understanding-first condition on students' ability to transfer fractions knowledge to novel task types. To this end, I used the Tetrad IV program[5] to search for models that are theoretically plausible and consistent with the data. Specifically, I used the GES algorithm in Tetrad IV along with background knowledge constraining the space of models searched (Chickering, 2002) to those that are theoretically tenable and compatible with the experimental design. In particular, I assumed that condition is exogenous and causally independent, that pretest is exogenous and causally independent from condition, and that the mediators are prior to the posttest. Fig. 37 illustrates the fully saturated model that would be compatible with these assumptions.

The qualitative causal structure of each of these linear structural equation models can be represented by a Directed Acyclic Graph (DAG). If two DAGs entail the same set of constraints

[5] 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).

on the observed covariance matrix,[6] then they are empirically indistinguishable. If the constraints considered are independence and conditional independence, which exhaust the constraints entailed by DAGs among multivariate normal varieties, then the equivalence class is called a *pattern* (Pearl, 2000; Spirtes et al., 2000). Instead of searching in the DAG space, the GES algorithm achieves significant efficiency by searching in pattern space. The algorithm is asymptotically reliable,[7] and outputs the *pattern* with the best BIC score.[8] The pattern identifies features of the causal structure that are distinguishable from the data and background knowledge, as well as those that are not. The algorithm's limits are primarily in its background assumptions involving the non-existence of unmeasured common causes and the parametric assumption that the causal dependencies can be modeled with linear functions.

I conducted several mediation analyses, testing the mediation of the effect of condition on the transfer posttest through accuracy on connection-fluency quiz items, efficiency on connection-fluency quiz items, error rates on fluency-building knowledge components, and error rates on sense-making knowledge components.

Mediation of condition effects through performance on connection-fluency quiz

Let us first consider the analysis that investigated whether the differences between conditions on *accuracy on connection-fluency* quiz items mediates the advantage of the understanding-first condition over the fluency-first condition on *accuracy* on the transfer posttest. Fig. 38 shows a model found by GES, with coefficient estimates included. The model fits the data well, (χ^2 = 2.43, df = 4, p = .66). Students in the understanding-first condition perform better than students in the fluency-first condition on accuracy on connection-fluency quiz at quiz time 2 (that is, after having received fluency-building support). Higher accuracy on the connection-fluency quiz at quiz time 2 increases accuracy on the transfer posttest, after controlling for accuracy on the transfer pretest.

[6] An example of a testable constraint is a vanishing partial correlation, e.g., $\rho_{XY.Z} = 0$.

[7] Provided the generating model satisfies the parametric assumptions of the algorithm, the probability that the output equivalence class contains the generating model converges to 1 in the limit as the data grows without bound. In simulation studies, the algorithm is quite accurate on small to moderate samples.

[8] All the DAGs represented by a pattern will have the same BIC score, so a pattern's BIC score is computed by taking an arbitrary DAG in its class and computing its BIC score.

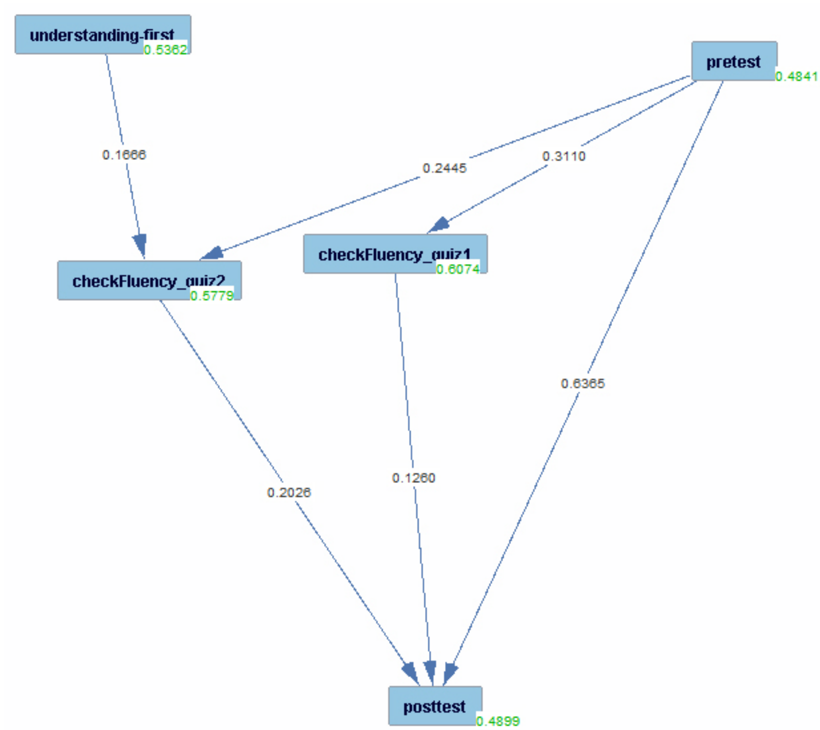


Fig. 38. Experiment 5: The model found by GES for the mediation hypothesis of the effect understanding-first condition on accuracy posttest transfer through accuracy on connection-fluency quiz items.

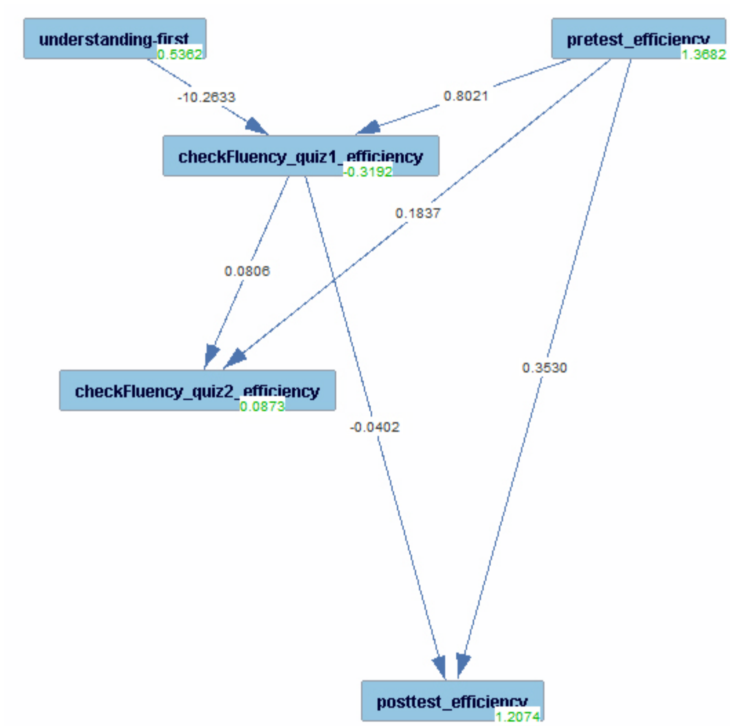


Fig. 39. Experiment 5: The model found by GES for the mediation hypothesis of the effect understanding-first condition on *efficiency* posttest transfer through *efficiency* on connection-fluency quiz items.

Next, I investigated whether the differences between conditions on *efficiency on connection-fluency* quiz items mediates effects on *efficiency* on the transfer posttest. Fig. 39 shows a model found by GES, with coefficient estimates included. The model fits the data well, ($\chi^2 = 3.38$, df = 4, $p = .50$). Students in the understanding-first condition perform worse than students in the fluency-first condition on efficiency on connection-fluency quiz at quiz time 1 (that is, before having received fluency-building support). Higher efficiency on the connection-fluency quiz at quiz time 1 slightly decreases efficiency on the transfer posttest, after controlling for efficiency on the transfer pretest. Thus, efficiency on connection-fluency at quiz time 1 mediates a very slight negative effect of the understanding-first condition on efficiency on the transfer posttest.

Taken together, the mediation analysis sheds light into how the differences between conditions on accuracy and efficiency on the connection-fluency quizzes are to be interpreted. The advantage of the understanding-first condition over the fluency-first condition on the transfer posttest is fully mediated by increased accuracy on the connection-fluency quiz at quiz time 2. Thus, *after* having received sense-making support, students benefit more from receiving fluency-building support than students who receive the same fluency-building problems *before* sense-making support. It is important to note that this advantage only plays out in accuracy measures, not in efficiency measures. Altogether, this finding is in line with the interpretation of the mediation analysis on error types in Experiment 4 (see section 4.4.3.2), that conceptual understanding of the connections between multiple graphical representations enables students to benefit from connectional fluency-building support. Students' ability to speedily solve connection-fluency problems (i.e., efficiently solving such problems) does not benefit from having previously received connectional sense-making support. However, students in the understanding-first condition make up for a disadvantage in efficiency that appears at quiz time 1 after having received connectional fluency-building support at quiz time 2. Their ability to efficiently transfer their knowledge about fractions to novel task types suffers only very slightly from this early disadvantage in efficiently solving connection-fluency problems at quiz time 1.

Mediation of condition effects through problem-solving behaviors

Further, I investigated whether the differences between conditions on *error rates on fluency-building knowledge components* (see Table 24) mediates the advantage of the understanding-first condition over the fluency-first condition on *accuracy* on the transfer posttest. Fig. 40 shows a model found by GES, with coefficient estimates included. The model fits the data well, (χ^2 = 4.58, df = 4, p = .33). Students in the understanding-first condition show lower error rates on equivalence fluency-building knowledge components than students in the fluency-first condition. Higher error rates on equivalence fluency-building knowledge components decrease accuracy on the transfer posttest. Thus, error rates on equivalence fluency-building knowledge components fully mediate a positive effect of the understanding-first condition on accuracy on the transfer posttest: students who received connectional sense-making support before connectional fluency-building support show lower error rates while working on equivalence fluency-building support problems, which accounts for their advantage on accuracy on the transfer posttest.

Next, I investigated whether the differences between conditions on *error rates on sense-making knowledge components* (see Table 24) mediates the advantage of the understanding-first condition over the fluency-first condition on *accuracy* on the transfer posttest. Fig. 41 shows a model found by GES, with coefficient estimates included. The model fits the data well, (χ^2 = 3.38, df = 3, p = .38). Students in the understanding-first condition show lower error rates on equivalence sense-making knowledge components and on comparison sense-making knowledge components than students in the fluency-first condition. Higher error rates on equivalence sense-making knowledge components also lead to higher error rates on comparison sense-making knowledge components. Higher error rates on comparison sense-making knowledge components decrease accuracy on the transfer posttest. Thus, error rates on sense-making knowledge components fully mediate a positive effect of the understanding-first condition on accuracy on the transfer posttest: students who did not receive connectional fluency-building support before working on sense-making support problems show lower error rates while working on sense-making support problems, which accounts for their advantage on accuracy on the transfer posttest.

Taken together, the mediation analysis provides additional support for the understanding-first hypothesis and, furthermore, yields insights into the mechanisms by which the understanding-first condition leads to higher accuracy on the transfer posttest. First, the mediation analysis supports the interpretation of the mediation analysis on error types in Experiment 4 (see section 4.4.3.2), that connectional understanding enables students to benefit from connectional fluency-building support. Connectional sense-making support reduces the error rate on equivalence flu-

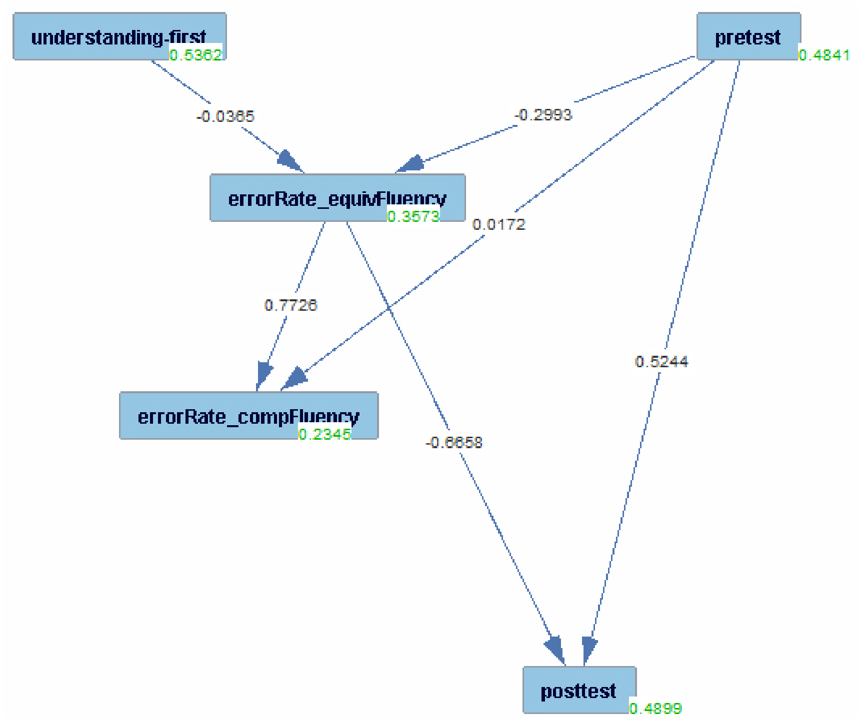


Fig. 40. Experiment 5: The model found by GES for the mediation hypothesis of the effect understanding-first condition on *accuracy* posttest transfer through *error rates on fluency-building* knowledge components.

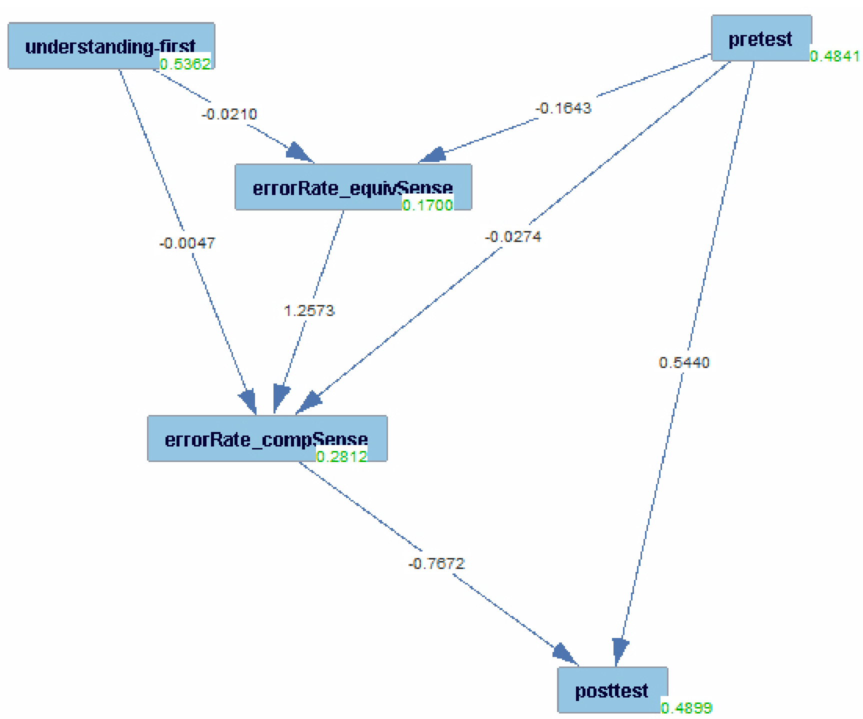


Fig. 41. Experiment 5: The model found by GES for the mediation hypothesis of the effect understanding-first condition on *accuracy* posttest transfer through *error rates on sense-making* knowledge components.

ency-building problems, which accounts for the higher accuracy of the understanding-first condition on the transfer posttest.

Second, the mediation analysis of error rates on connectional sense-making support complements this finding by demonstrating potential costs of receiving connectional fluency-building support before connectional sense-making support. Students in the understanding-first condition show lower error rates than students in the fluency-first condition while working on comparison sense-making problems. In other words, students in the fluency-first condition (i.e., students who received connectional fluency-building support *before* working on connectional sense-making problems) show *higher* error rates on sense-making problems, which accounts for their *lower accuracy* on the transfer posttest. Thus, connectional fluency-building support decreases students' ability to benefit from connectional sense-making support, thereby hampering students' acquisition of robust domain knowledge that can transfer to novel task types.

4.5.4 Discussion

Prior research shows that both sense-making processes and fluency-building processes play an important role in connection making: both learning processes need to be supported in order for students' robust learning of domain knowledge to benefit from multiple graphical representations (Rau, Aleven et al., 2012). The results from Experiment 5 shed light on the question of *how* these learning processes interact. I contrasted two competing hypotheses. On the one hand, the *understanding-first hypothesis* posits that connectional sense-making support enhances students' benefit from subsequent connectional fluency-building support by helping them focus on conceptually relevant aspects of graphical representations. According to the *fluency-first hypothesis*, on the other hand, connectional fluency-building support enhances students' benefit from subsequent connectional sense-making support by freeing cognitive resources that students can invest in sense-making processes to develop connectional understanding. I investigated contrasting predictions made by each of these hypotheses for both learning outcomes and process-level measures.

Altogether, the results from Experiment 5 are in line with the understanding-first hypothesis, but not with the fluency-first hypothesis. Students in the understanding-first condition outperformed students in the fluency-first condition on accuracy of connection-fluency at quiz time 2, which accounts for the advantage of the understanding-first condition on accuracy on the transfer posttest. Further, even though students in the understanding-first condition showed lower efficiency on connection-fluency quiz items, there were no differences between conditions on students' efficiency on the transfer posttest. Rather, as the mediation analysis of student's efficiency on the connection-fluency quiz items shows (see Fig. 39), students' lower efficiency on connection-fluency quiz items was associated with higher efficiency on the transfer posttest. Thus, with regard to learning outcomes, the findings are in line with the understanding-first hypothesis, but not with the fluency-first hypothesis. Receiving connectional sense-making support *before* connectional fluency-building support enables students to acquire robust, accurate knowledge of fractions that transfers to novel task types.

The analysis of process-level measures provides insights into the mechanisms underlying the advantage of the understanding-first condition on students' accuracy on the transfer posttest. There are several aspects that explain the advantage of providing connectional sense-making support before connectional fluency-building support. First, as the analysis of problem-solving behaviors shows, connectional understanding enables students to benefit from connectional fluency-building support (on the earlier equivalence problems), whereas connectional fluency hampers students' benefit from connectional sense-making support (on the later sense-making problems). The mediation analysis demonstrates that these differences in students' problem-solving behaviors statistically explain the advantage of the understanding-first condition on transfer accuracy. The notion that connectional understanding benefits students' learning from connectional fluency-building support was anticipated and is in line with the understanding-first hypothesis. The finding that connectional fluency-building support harms students' benefit from subsequent connectional sense-making support (as assessed by error rates on equivalence-fluency problems), however, is unexpected. It may be that connectional fluency-building support "primes" students to rely on perceptual characteristics rather than to conceptually think about connections, making them more "careless" as they intuitively go about solving sense-making support problems. This interpretation is in line with the concern expressed by Bieda and Nathan (2009) that students who are overly influenced by the perceptual properties of a representation may not pay attention to the conceptually relevant aspects of a representation, which is crucial to their learning of domain knowledge.

Second, the analysis of the retrospective interviews collected in phase 1 of the experiment shows that the understanding-first condition engages in more conceptual reasoning about fractions than the fluency-first condition (as assessed posthoc by retrospective interview data). At the same time, connectional fluency-building support elicits significantly more connection-making utterances than connectional sense-making support, although the number of fluency-connections did not differ between conditions. A reasonable interpretation of these findings may be that the combination of connectional sense-making and fluency-building support is necessary because only connectional fluency-building support promotes explicit connection making between multiple graphical representations. However, only when students receive connectional sense-making support *before* fluency-building support can they benefit more from these connections, by conceptually reasoning about the connections rather than being overly influenced by the perceptual properties of the graphical representations.

It is critical to note that the advantage of the understanding-first condition over the fluency-first condition plays out only on accuracy on transfer, but not on efficiency. Both conditions perform equally well when considering efficiency measures on the transfer posttest. How might one explain the lack of differences between conditions on efficiency on the transfer posttest? This finding makes sense when considering the earlier stated interpretation that the fluency-first condition is “primed” to make use of the perceptual characteristics of graphical representations to solve transfer problems. Perhaps the fluency-first condition focuses on becoming more efficient at solving transfer problems rather than solving them more accurately. This early focus on efficiency rather than on accuracy might come at the expense of lower accuracy in transferring fractions knowledge to novel problems. However, for the understanding-first condition, the focus on improving on accuracy rather than on efficiency appears not to come at the expense of lower efficiency in solving fractions problems: there are no differences between the two conditions on the accuracy measure of the transfer posttest.

It is important to point out that these findings leave some room for doubt: differences were found on only some of the dependent measures; the effects of sense-making support and fluency-building support appears to differ by topic. Furthermore, the effects of condition on accuracy of transfer were only marginally significant. Further research is needed to support the conclusion that connectional sense-making support should be provided before fluency-building support, which mechanisms account for the advantage of the understanding-first condition over the fluency-first condition, and to investigate which properties of topics (e.g., equivalence versus comparison) or the time during the learning process under consideration (e.g., earlier versus later) enhance or diminish this effect. Especially with respect to the proposed mechanisms that underlie the advantage of the understanding-first condition, more research is needed. For instance, eye-tracking data can be used to investigate whether connectional understanding, as described by the understanding-first hypothesis, indeed helps students direct their attention to relevant aspects of graphical representations in while they work on fluency-building problems. Think-aloud data collected while students work on the tutor problems might also yield insights into what aspects of graphical representations pay attention to, and how these processes differ between conditions. A larger sample size would provide more statistical power to compare alternative, and more exhaustive mediation models that take into account a larger number of dependent measures than the separate mediation models I computed for the given (relatively small) data set. In spite of this caveats, it is striking that the effects found on a number of learning outcomes and process-level measures are in line with the understanding-first hypothesis, whereas no single effect was in line with the fluency-first hypothesis. Therefore, it seems reasonable to conclude that, to promote accuracy of transfer of domain knowledge, connectional sense-making processes should be supported before connectional fluency-building processes.

In summary, the results from Experiment 5 support and extend the understanding-first hypothesis: connectional understanding not only *enhances*, but *is necessary* for students' benefit from connectional fluency-building support because it enables students to relate connections between multiple graphical representations to conceptual knowledge about fractions. Whether or not these conclusions hold for other domains than fractions learning remains to be empirically tested. Based on Experiment 5, I conclude that instructional designers of multi-representational learning materials should provide students with connectional sense-making support before connectional fluency-building support, in particular when their goal is to promote not only efficiency in problem solving, but also accuracy in problem solving.

Experiment 5 extends my theoretical framework for learning with multiple graphical representations by clarifying how two of the proposed learning processes interact, based on the finding from Experiment 4. Connectional fluency-building processes appear to, at least to some extent, build on connectional understanding. On the other hand, connectional sense-making processes might be hindered by connectional fluency. If perceptual chunking is the mechanism that connectional fluency-building support enhances, the acquisition of more coarse-grained perceptual chunks might hinder students' ability to acquire more fine-grained chunks, as supported by connectional sense-making processes that require students to reason separately about constituent knowledge components. Instead, the availability of a perceptual chunk might "bypass" the learning of a connectional sense-making knowledge component, because a student can solve the problem based on the perceptual chunk, without having connectional knowledge components that correspond to separate aspects of graphical representations, such as numerator and denominator. Yet, as the results from the learning outcomes on accuracy in transferring fractions knowledge suggest, the acquisition of connectional sense-making knowledge components facilitate learning of robust domain knowledge. This interpretation of the results from Experiment 5 with respect to my theoretical framework is speculative, given that I did not include a knowledge component model that assesses knowledge components for connectional understanding and for connectional fluency separately. It would be interesting to develop such a knowledge component model and test these predictions. For example, does the acquisition of connectional fluency (e.g., of a knowledge component "connect-circle-nl") hinder the acquisition of connectional understanding (e.g., of the knowledge components "connect-numerator-circle-nl" and "connect-denominator-circle-nl")? Furthermore, these questions could be extended to representational sense-making and fluency-building processes as well: does the acquisition of representational fluency (e.g., of a knowledge component "ident-fract-circle") hinder the acquisition of representational understanding (e.g., of knowledge components "ident-numerator-circle" and "ident-denominator-circle")?

Although, as mentioned, speculative in certain aspects, these reflections illustrate how iterating between the theoretical framework for learning with multiple graphical representations, experimental studies that test predictions made by the framework (such as Experiment 4), experimental studies to test hypotheses that follow from resulting findings (such as Experiment 5), and the interpretation of these results in terms of the theoretical framework (as, for example, described in this section), can stimulate further hypotheses and, potentially, empirical research that in turn, integrates learning outcome measures and assessments of learning processes.

5 Conclusion

Taken together, my dissertation work comprises a sequence of experimental studies that focus on learning with multiple graphical representations and the development of an intelligent tutoring system for fractions learning. This research reflects two perspectives of my research: a learning sciences perspective and an educational technology perspective. At the heart of both contributions lies a set of instructional design recommendations for the effective use of multiple graphical representations. These design principles are the outcome of a series of classroom experiments. In addition to providing practical guidance for developers of multi-representational learning materials, these design principles extend the literature on learning with multiple representations. The results from my experiments can be interpreted with respect to a theoretical framework that I iteratively developed throughout the sequence of experiments and tutor development process. Furthermore, the experiments serve to iteratively improve the Fractions Tutor, which has repeatedly been shown to lead to robust learning of fractions knowledge. Finally, I describe a novel methodology for resolving conflicts between stakeholder goals, which has fundamentally shaped the design of the Fractions Tutor.

In this section, I first review the contributions of my empirical work to the learning sciences perspective of my dissertation work, which include both the instructional design recommendations, contributions to existing theoretical perspectives on learning with multiple representations, and a novel theoretical framework for learning with multiple graphical representations. I then discuss my contributions from an educational technology perspective on my research. After critically discussing limitations of my work, reflecting on scenarios to which I expect my results to generalize, and describing possible future directions, I end by discussing the integration of the learning sciences and educational technology perspectives of my work through the use of a multi-methods approach.

5.1 Learning sciences perspective

There are two aspects to the learning sciences perspective of my work: providing practical recommendations for the effective use of multiple graphical representations in instructional materials, and extending existing theory about how multiple representations influence students' learning of domain knowledge. Finally, I discuss my findings from the perspective of my theoretical framework for learning with multiple graphical representations.

5.1.1 Instructional design principles for the use of multiple graphical representations

Prior research on multiple representations has mostly taken a symbol-systems approach and investigated learning with textual descriptions accompanied by only one additional graphical representation (e.g., Ainsworth & Loizou, 2003; Baetge & Seufert, 2010; Bodemer et al., 2005; Butcher & Aleven, 2007; Kuehl et al., 2010; Magner et al., 2010; Rasch & Schnotz, 2009; Suthers et al., 2008). This focus stands in contrast with the common use of *multiple graphical* representations (e.g., Arcavi, 2003; Cook et al., 2007; Kordaki, 2010; Kozma et al., 2000; Urban-Woldron, 2009; Walkington et al., 2011), which are provided in addition to textual and symbolic representations in many educational materials. My research addresses the gap between the common symbol-systems approach in prior research on multiple representations and the resulting lack of guidance for the design of more complex materials that include multiple graphical representations. In doing so, I provide a set of instructional design principles that result from my experimental studies.

First, multiple graphical representations do indeed have the potential to enhance students' learning of domain knowledge, compared to a single graphical representation, as demonstrated by Experiments 1, 3, and 4. These experiments have generally found advantages of multiple graphical representations on learning of robust conceptual knowledge. Thus, the first design principle, which provides the foundation for my thesis work, is:

Principle 1: Use multiple graphical representations (with appropriate instructional support) to enhance robust learning of domain knowledge.

The qualification "with appropriate instructional support" describes the fact that multiple graphical representations do not automatically enhance students' learning, as my experimental studies repeatedly illustrate. In fact, Experiment 1 shows that multiple graphical representations only lead to better learning than a single graphical representation if they are provided along with reflection prompts that encourage students to self-explain the relation between each graphical representation and the corresponding symbolic representation. The second design principle therefore is:

Principle 2: Use reflection prompts to support students in relating graphical representations to the key concepts they depict.

Experiments 2 and 3 investigate the effects of interleaved practice with task types and graphical representations. Experiment 2 shows that interleaving task types (while blocking graphical representations) enhances students' learning of representational knowledge more so than interleaving graphical representations (while blocking task types). Since differences between task types are more salient than differences between graphical representations, I conclude that interleaving learning tasks along the dimension of greatest variability leads to more robust learning of robust conceptual understanding of graphical representations, although it remains an open question whether this interpretation holds as a general principle. Consequently, the third design principle is:

Principle 3: Interleave learning tasks along the dimension of task types.

While employing a moderately interleaved schedule of task types consistently across conditions, Experiment 3 shows that, interleaving of graphical representations also enhances students' learning of conceptual knowledge. The fourth design principle therefore states:

Principle 4: In addition to (moderately) interleaving task types, also interleave graphical representations.

Experiment 4 investigates how best to support students in making connections between multiple graphical representations. While neither connectional sense-making support nor connectional fluency-building support alone were effective, they enhanced students' robust learning of conceptual knowledge when combined. Thus, the fifth design principle is:

Principle 5: Combine connectional sense-making support and connectional fluency-building support, rather than providing either type of support alone.

With respect to how best to implement connectional sense-making support, Experiment 4 shows that students need to become active in establishing the connections between graphical representations themselves, rather than being provided with the connections semi-automatically by the system. Consequently, the sixth design principle is:

Principle 6: Implement connectional sense-making support in a way that requires students to actively generate connections, rather than passively observing them.

Finally, Experiment 5 investigates how connectional sense-making support and connectional fluency-building support interact, and which type of support should be provided first. The results show that in order to support students' accuracy in transferring domain knowledge to novel tasks, connectional sense-making support should be provided before fluency-building support. Therefore, the seventh design principle is:

Principle 7: Provide connectional sense-making support before connectional fluency-building support.

Needless to say, it is open to further investigation whether these design principles hold across other domains and other means of implementing these recommendations than was done in my experimental studies. I am confident that the critical discussion of the limitations of each experiment (see section 4) has made it evident that I view these design principles not as the final word of wisdom, but rather as the outcome of a specific set of studies that can serve as working hypotheses for future research. I critically discuss the limitations of my work later in this section (see section 5.3). However, given that my thesis work is, to the best of my knowledge, the first comprehensive set of experimental studies to investigate how best to implement *multiple graphical* representations that use *the same* symbol system, I hope that in spite of the limited nature of this work, developers of instructional materials will see the merit of these instructional design principles.

5.1.2 *Extending prior theoretical perspectives on learning with multiple representations*

Taken together, my research shows that the advantage of multiple representations is not limited to learning with representations from *different* symbol systems, but applies also to the common scenario that *multiple graphical* representations, which use *the same* symbol system, are provided along with textual descriptions and symbolic representations.

Experiment 1 provides the foundation for my dissertation research. In particular, the merit of providing students with multiple graphical representations using the same symbol system, as is common practice in many educational materials, remained to be experimentally established –

due to the symbol-systems approach taken by most prior research on learning with multiple representations. As argued in section 2.1, existing theoretical frameworks take a symbol-systems approach and therefore cannot explain why multiple graphical representations might enhance learning. By contrast, they might lead to the prediction that multiple graphical representations do not benefit learning more than a single graphical representation because they do not require students to integrate information across different symbol systems (Schnotz & Bannert, 2003), and may even harm learning by resulting in cognitive overload in the pictorial part of working memory (Mayer, 2003; Mayer 2005).

The results from Experiment 1 show that multiple graphical representations lead to better learning than a single graphical representation, provided that students are prompted to reflect on the relation between graphical and symbolic representations. Experiment 3 provides further support for the notion that multiple graphical representations can enhance robust domain knowledge compared to a single graphical representation. The finding that only students who worked with multiple graphical representations showed significant learning gains, whereas a variety of single-representation conditions did not show learning gains, extends Experiment 1 by demonstrating that the advantage of multiple graphical representations does not depend on the specific graphical representation used in the single-representation control condition. Also Experiment 4 shows a significant advantage of multiple graphical representations over a single graphical representation, provided that students receive both sense-making and fluency-building support for connection making. In general, Experiments 1, 3, and 4 found advantages of multiple graphical representations on tests that assessed conceptual understanding rather than procedural knowledge, demonstrating that the merit of using multiple graphical representations lies in enhancing deep processing of the conceptual aspects of the learning material, and not in enhancing the ability to carry out computational operations. Taken together, my thesis work extends prior research that was mostly conducted under the symbol-systems assumption. My research demonstrates that multiple representations that use *the same* symbol system (e.g., *multiple graphical* representations) do indeed have the potential to enhance learning of robust domain knowledge. In other words, multiple representations of the same symbol system (i.e., *multiple graphical* representations) can lead to better learning than a single representation of that symbol system (i.e., a *single* *graphical* representation). Consequently, the advantage of multiple representations is not limited to representations of different symbol systems.

I attribute this advantage of multiple graphical representations to the purpose with which multiple graphical representations are being used in instructional materials, as detailed in sections 1.1 and 3.1. As in many STEM domains, fractions instruction uses different graphical representations to emphasize specific conceptual aspects of the domain (Charalambous & Pitta-Pantazi, 2007). To deeply understand the conceptual aspects of fractions, students need to integrate the different conceptual views depicted by the different graphical representations. This reflection illustrates that my research does not contradict the notion that multiple representations are beneficial because they enhance deeper processing due to the integration of information across symbol systems, as expressed in Schnotz and Bannert's (2003) framework for learning with multiple representations. Rather, my research extends this framework: what is crucial to the positive effect of multiple representations on students' learning is the potential to encourage students' engagement in deep, conceptual processing of the structural elements that constitute the information depicted in different representations. This type of integration process does not have to occur across different symbol systems – it can also occur between multiple representations that use the same symbol system.

Yet, multiple graphical representations do not automatically result in more robust learning than a single graphical representation, as the results from Experiments 1-5 demonstrate. Rather, their success may depend on (or at least be enhanced by) the implementation of multiple graphical representations in accordance to the instructional design principles summarized above (see section 5.1.1). The finding that the benefit of multiple graphical representations may depend on how they are implemented is in line with a number of studies that show that multiple representations (using different symbol systems) do not automatically result in better learning than text alone (Ainsworth, 2006; Ainsworth, Bibby, & Wood, 1998; de Jong et al., 1998; Kim et al., 2013; Tsui & Treagust, 2013). These results also further highlight the importance of establishing principles for instructional support for learning with multiple graphical representations (see section 5.1.1).

The finding in Experiment 1, that reflection prompts enhance students' benefit from multiple graphical representations, is in line with a vast literature on learning with multiple representations (that use different symbol systems), which show that (1) the advantage of learning with text and graphics (as opposed to text alone) may stem from multiple representations enhancing students' tendency to self-explain (Ainsworth & Loizou, 2003), and that (2) interventions that promote self-explanation activities can enhance students' benefits from multiple representations (Berthold et al., 2008; Berthold & Renkl, 2009; Zhang & Linn, 2011). Yet, Experiment 1 is, to the best of my knowledge, the first study to systematically investigate the interaction between the effect of reflection prompts and the effect of multiple representations. Furthermore, Experiment 1 extends these findings to learning with multiple graphical representations (that use the same symbol system). Experiment 1 shows that students' benefit from multiple graphical representations is not merely enhanced by reflection prompts, but *depends* on receiving such support. The reflection prompts in the Fractions Tutor are designed to support students in self-explaining the relation of each component of a given graphical representation (e.g., the number of colored sections in a circle) to the corresponding knowledge component as depicted symbolically (e.g., the numerator of a fraction). These reflection prompts used in Experiment 1 are thus equivalent to the type of connection-making support often used in research on connection making between multiple representations that use different symbol systems (Schwonke, Berthold et al., 2009; Seufert, 2003). Therefore, one might also interpret the finding from Experiment 1 as showing that connection making between representations of different symbol systems (i.e., between the graphical and symbolic representation) is necessary for students to benefit from multiple graphical representations that use the same symbol system.

Experiment 2 contrasts the effects of interleaving graphical representations (while blocking task types) and interleaving task types (while blocking graphical representations) on students' learning. Experiment 2 built on the findings of Experiment 1 in that both conditions in Experiment 2 contained the same reflection prompts used in Experiment 1. Results show that interleaving task types (while blocking graphical representations) leads to higher accuracy and efficiency in using conceptual representational knowledge to solve fractions problems than interleaving graphical representations (while blocking task types). This finding extends the literature on the contextual interference effect (de Croock et al., 1998; van Merriënboer et al., 2002), by showing that it matters which dimension of learning tasks are interleaved. I argue that task types constitute the more variable dimension, compared to graphical representations, on which learning tasks can vary, because the differences between task types are more salient than the differences between graphical representations. Furthermore, interleaving learning tasks along the dimension of greatest variability might be most effective because it increases the need for students to frequently reactivate knowledge (de Croock et al., 1998; Lee & Magill, 1983, 1985) as well as the number of opportunities for abstracting across consecutively presented learning tasks (de Croock et al., 1998; Shea & Morgan, 1979).

The finding that practice schedules have an impact on students' learning with multiple graphical representations is also interesting in the light of prior research on multiple representations. This prior research has mostly investigated explicit types of instructional support for students' learning with text and graphic, for instance, by providing active means of referencing (Bodemer & Faust, 2006; Bodemer et al., 2005; Bodemer et al., 2004), by providing help features (Brünken, Seufert, & Zander, 2005; Seufert, 2003), instructional aids (Seufert, 2003), prompts for self-explanation activities (Butcher & Aleven, 2007), or trainings (Schwonke et al., 2008; Wong, Lawson, & Keeves, 2002). Experiment 2 demonstrates that subtle variations, such as the sequence in which graphical representations are provided across different task types, have an impact on which aspects of the learning tasks students strengthen and abstract across. Thereby, the choice of practice schedule implicitly affects students benefit from multiple graphical representations, that is, without explicit support.

Experiment 3 focuses on the effects of interleaving graphical representations while using a consistent schedule of task types across all conditions. Building on the finding of Experiment 2, Experiment 3 investigates whether *in addition to* moderately interleaving task types, one should also interleave graphical representations. The results from an analysis of learning outcomes and tutor log data provide moderate evidence that interleaving graphical representations leads to more robust learning of domain knowledge than blocking graphical representations. This finding extends prior research on multiple representations by showing that the sequence in which representations are provided affects students' learning. Furthermore, Experiment 3 extends research on practice schedules, which has mostly focused on the sequence of different task types. Experiment 3 shows that the advantage of interleaving task types generalizes to interleaving graphical representations.

A small-scale think-aloud study sheds light into the question whether abstraction across graphical representation or repeated reactivation of representation-specific knowledge might explain the advantage of interleaved practice. The results from the think-aloud study stand in contrast to one possible explanation for the advantage of interleaving graphical representations, namely, that students spontaneously abstract across graphical representations by making explicit connections between consecutively presented representations. In the light of this finding, it seems more likely that interleaved practice with multiple graphical representations requires students to repeatedly reactivate representation-specific knowledge in the interleaved condition, thereby strengthening their knowledge about each individual graphical representations, which increases the likelihood that they can retrieve it faster and with less effort in the future.

An analysis of the tutor log data by the means of Bayesian Knowledge Tracing demonstrates higher learning rates for students in the interleaved condition, compared to students in the blocked condition. In applying Bayesian Knowledge Tracing to research on interleaved practice, Experiment 3 further extends prior research on practice schedule effects that has failed (just like Experiment 3) to show an advantage of interleaved practice over blocked practice using *raw performance measures* obtained during the acquisition phase. By modeling a latent variable of students' learning, using Bayesian Knowledge Tracing, I provide evidence of an advantage of the interleaved condition over the blocked condition, even during the acquisition phase. Experiment 3 thus demonstrates that latent variable modeling, rather than raw performance measures, are a suitable metric to studying the effects of practice schedules on learning during the acquisition phase.

Experiment 4 investigates the complementary effects of supporting students in acquiring connectional understanding and of supporting them in acquiring connectional fluency. The findings extend prior research, which has focused only on either learning process alone: either on connectional sense-making processes (Bodemer & Faust, 2006; Bodemer et al., 2004; Plötzner et al., 2001; Schwonke et al., 2008; Seufert, 2003; van der Meij & de Jong, 2006), or on connectional fluency-building processes (Kellman & Garrigan, 2009; Kellman et al., 2008; Kellman et al., 2009; Massey et al., in press). Furthermore, Experiment 4 extends this research by focusing on connection making between multiple graphical representations, that is, on representations using *the same* symbol system, whereas prior research has focused on connection making between representations using *different* symbol systems, such as text and graphical representation (Bodemer & Faust, 2006; Bodemer et al., 2004; Plötzner et al., 2001; Seufert, 2003), between symbolic representation and graphical representation (van der Meij & de Jong, 2006), or between text, symbolic representations, and graphical representation (Kellman & Garrigan, 2009; Kellman et al., 2008; Kellman et al., 2009; Massey et al., in press). The results from Experiment 4 indicate that students' benefit from multiple graphical representations depends on combining support for both connectional sense-making processes and fluency-building processes: only students who received a combination of both types of support significantly outperformed students in a single-representation control condition on measures of robust conceptual knowledge of fractions.

Furthermore, Experiment 4 investigates the role of system-based support for connectional sense-making processes. While some research suggests that automated support for connectional sense-making processes can enhance learning (van der Meij & de Jong, 2006), other research shows that students need to actively engage in connectional sense-making processes (e.g., Bodemer & Faust, 2006; Bodemer et al., 2004; Gutwill et al., 1999). Results from Experiment 4 support the latter notion and show that worked examples, which require students to actively generate the connections themselves rather than relying on support that is provided semi-automatically by the system, is the more effective type of connectional sense-making support. Worked examples have been shown to be effective in supporting sense-making processes in a variety of domains (e.g., Berthold et al., 2008; Große & Renkl, 2007; Kopp et al., 2008; McLaren et al., 2008; Nokes & VanLehn, 2008; Renkl, 2005; Schwonke et al., 2009). Prior research also established that multiple representations (i.e., text and graphical representation) can make worked examples more effective (e.g., Berthold et al., 2008; Berthold & Renkl, 2009; Schwonke, Berthold et al., 2009). But prior research has not investigated the effect of worked examples as a means to support connection making between multiple graphical representations. Experiment 4 extends this prior research by showing that worked examples can increase the effectiveness of multiple graphical representations, when used to support connectional sense-making processes (and in conjunction with support for connectional fluency-building processes).

An analysis of the Experiment 4 log data by the means of causal path modeling provides tentative insights into how connectional sense-making support and fluency-building support interact. Students who receive sense-making support in addition to fluency-building support make fewer errors on fluency-building problems compared to students who receive no sense-making support. The reduction in errors made on the fluency-building problems accounts for the advantage of receiving a combination of both types of support (compared to receiving fluency-building support only) on the conceptual knowledge posttest. By contrast, receiving fluency-building support does not influence the frequency of errors students make on the sense-making problems. The analysis of the tutor log data leads to a hypothesis that can be tested empirically; namely, the understanding-first hypothesis, that the combination of connectional sense-making support and fluency-building support promotes students' learning *because* sense-making support enables students to benefit from connectional fluency-building support. Thus, the combination of sense-making support and fluency-building support should be most effective in supporting students' robust learning of domain knowledge if support for connectional sense-making processes is provided *before* support for connectional fluency-building processes.

Experiment 5 was designed to contrast the understanding-first hypothesis and the fluency-first hypothesis. According to the understanding-first hypothesis, connectional understanding equips students with prerequisite knowledge that allows them to attend to relevant aspects of graphical representations while working on connectional fluency-building problems. Not having connectional understanding might leave students at a loss of what aspects of different graphical representations are structurally equivalent, leading to inefficient learning strategies, which diminishes their benefit from connectional fluency-building support. In other words, the understanding-first condition posits that connectional fluency cannot be acquired without prerequisite connectional understanding. By contrast, the fluency-first condition holds that connectional fluency-building support equips students with (somewhat intuitive) perceptual knowledge about correspondences between different graphical representations. Students who have this type of perceptual fluency in making connections might experience lower cognitive load while making sense of connections between graphical representations, and might therefore be expected to benefit more from connectional sense-making support, compared to students who have not previously acquired connectional fluency.

Experiment 5 tested specific predictions by both hypotheses about students' learning outcomes on reproduction of connection-making tasks and their ability to transfer knowledge about the domain to novel task types, on problem-solving behaviors, and conceptual reasoning. Taken together, the results from Experiment 5 are in line with the understanding-first hypothesis but not the fluency-first hypothesis, at least when considering measures of accuracy. Students who receive connectional sense-making support before connectional fluency-building support outperform students who receive connectional fluency-building support before connectional sense-making support on measures accuracy of reproduction on connection-fluency tasks and of a transfer test of fractions knowledge. However, when considering measures of efficiency (i.e., the ability to speedily and accurately solve these tasks), the sequence in which connectional sense-making and fluency-building support are provided does not have a substantial effect. Providing connectional fluency-building support before connectional sense-making support appears to lead students to prioritize on becoming more efficient in solving fractions tasks, which comes at the expense of becoming more accurate. By contrast, providing connectional sense-making support before connectional fluency-building support leads to marginally higher accuracy on the transfer posttest without diminishing students' development of efficiency in solving transfer problems.

The analysis of process-level measures sheds some light into these findings. On the one hand, connectional sense-making support *increases students' benefit* from connectional fluency-building support by (although slightly) reducing errors students make on connectional fluency-building problems presented earlier in the learning process (i.e., on equivalence-fluency problems). Similarly, providing connectional sense-making support before connectional fluency-building support leads to increased conceptual reasoning about fractions, as assessed posthoc by retrospective interviews. This finding is in line with the results from the causal path analysis in Experiment 4. On the other hand, connectional fluency-building support appears to *diminish students' benefit* from connectional sense-making support by increasing errors students make on sense-making problems provided later in the learning process (i.e., on comparison-sense problems). This finding extends the results from the causal path analysis in Experiment 4. Finally, connectional fluency-building support may be necessary to support connection making because it elicits more explicit connection-making utterances than sense-making support does, as assessed by retrospective interviews. This finding might provide some clarification about the interaction effect found in Experiment 4 by demonstrating why the combination of both types of support are needed to enhance students' benefit from multiple graphical representations: while connectional sense-making support enables students to benefit from connectional fluency-building support, only connectional fluency-building support leads to explicit and consciously accessible connection-making processes. It is important to point out that these interpretations about the *mechanisms* that underlie the advantage of the understanding-first condition over the fluency-first condition are highly speculative, in particular since the results on process-level measures were not unanimous. In particular, it appears that the effects of condition on students' problem-solving behaviors depended either on the type of topic (equivalence or comparison) or the time at which they were presented (earlier or later in the learning process). Since the topics in Experiment 5 were presented in a fixed rather than counterbalanced sequence, the relation between the topic and the time during the learning period at which these topics were presented remains unclear. More research is needed to investigate whether the effects of connectional sense-making support and connectional fluency-building support and students' benefit from these respective types of support hold across other topics than the ones tested, and across longer learning periods. Yet, even though the mechanism might remain somewhat unclear, the overall conclusion from Experiment 5 is that connectional sense-making support should be provided before connectional fluency-building support if the goal is to enhance students' accuracy in transferring domain knowledge to novel tasks.

Experiment 5 supports the somewhat implicitly held notion by many math education standards, which state an expectation for understanding of representations before the ability to efficiently work with them (e.g., NCTM, 2010). However, the findings extend Kellman and colleagues' (2009, in press) work, which is based on an implicit assumption that fluency in making connections can be acquired independently of connectional understanding. If connectional fluency can indeed be learned independently from connectional understanding, it is believed to reduce cognitive load while students engage in connectional sense-making processes later on. However, rather than reducing cognitive load, early connectional fluency-building support may lock students into a mode of learning that overly emphasizes perceptual properties of representations (Bieda & Nathan, 2009), which distracts students' attention from conceptual processing. Instead, students might become more careless at solving fractions problems, evidenced by their gains on the efficiency measure used in Experiment 5, and their lack of gains on the accuracy measure. Note, however, that the findings from Experiment 5 do not contradict Kellman and colleagues' work, since they never compared different sequences of fluency-building and sense-making support. However, it is important to note that Experiment 5 did not assess students' cognitive load during the learning process, such that this interpretation of the findings remain somewhat speculative and should be tested in future research.

In summary, my empirical work comprises a sequence of five experimental studies all of which extend the literature on learning with multiple representations that was mostly conducted under the symbol-systems assumption that multiple representations enhance students' learning by virtue of using different symbol systems. My experiments show that multiple graphical representations can enhance learning even though they use the same symbol system. However, to enhance their benefit from multiple graphical representations, students need to be supported in relating graphical representations to the symbolic notation of fractions, and to make sense of and become fluent in making connections between the different graphical representations. Moreover, interleaved practice with multiple graphical representations over an interleaved sequence of task types further enhances students' learning with multiple graphical representations. In addition to extending prior research on learning with multiple representations, my research integrates other literatures that have been considered separately: it integrates research on interleaved practice with the literature on learning with multiple representations, and it combines research on sense-making processes and fluency-building processes.

5.1.3 A new theoretical framework for learning with multiple graphical representations

As described in section 2.2, another contribution of my thesis is a theoretical framework that describes processes involved in learning with multiple graphical representations. The central claim of the framework is that instructional support for these hypothesized learning processes will increase students' learning of domain knowledge from working with multiple graphical representations. Since I developed this theoretical framework in iteration with my experimental studies, the results from my experiments constitute by no means an evaluation of the theoretical framework. Yet, the theoretical framework provides an interesting lens through which one might interpret some of the findings of my experimental studies – a lens that I hope will stimulate future research.

Experiment 1 shows that reflection prompts designed to help students self-explain the relation of each graphical representation to the corresponding symbolic notation based on knowledge components of numerator and denominator are necessary for students to benefit from multiple graphical representations. I argued that the learning processes that these reflection prompts support are most likely representational sense-making processes: reflection prompts are designed to help students make sense of each component of an individual graphical representation (e.g., number of shaded pieces in a circle) by relating it to the conceptual aspect of fractions it represents (e.g., the numerator of a fraction). Although Experiment 1 did not assess direct measures of sense-making processes (for example, through think-aloud protocols or interviews), it might be possible to interpret the results as suggesting that representational sense-making support (in the form of reflection prompts) enable students' to acquire robust domain knowledge from working with multiple graphical representations.

Experiments 2 and 3 show that interleaved practice with task types and graphical representations enhances students' acquisition of robust domain knowledge. As I described in the discussion of each experiment in section 4, it is unclear which of the proposed learning processes interleaved practice with task types and graphical representations support. The findings from the think-aloud study that was carried out as part of Experiment 3 suggest that interleaved practice with graphical representations does not support connection-making processes. Rather, interleaved practice might support either representational sense-making processes or representational fluency-building processes. I argue that interleaving task types (while blocking graphical representations) promotes representational sense-making processes by allowing students to abstract across different applications of the same graphical representation used over a sequence of task types, and to reactivate task-specific knowledge about one given graphical representation every time students switch between different task types. However, I hypothesize that interleaving graphical representations support representational fluency-building processes by allowing students to frequently reactivate representation-specific knowledge, thereby strengthening their knowledge and increasing the chance that students can fast and effortlessly recall that knowledge later on, which is an aspect of representational fluency as described in section 2.2. The finding from the Bayesian Knowledge Tracing analysis, that students in the interleaved condition show higher learning rates than students in the blocked condition, is in line with the interpretation that students become more fluent at using graphical representations to solve fractions problems – although alternative interpretations are possible. As these considerations are highly speculative and many alternative interpretations are possible (as described in sections 4.2.4 and 4.3.4), I would like to present these interpretations merely as careful hypotheses to be confirmed or rejected by future research.

Of all experiments, the relation between the theoretical framework and Experiment 4 is the strongest, since Experiment 4 tests a prediction that follows directly from the central claim of the theoretical framework: that support for connectional sense-making processes and connectional fluency-building should enhance students' learning of domain knowledge from multiple graphical representations. Experiment 4 shows indeed that providing both types of support leads to better learning of robust conceptual knowledge about fractions than either type of support alone. Experiment 4 even shows that only when both types of support are provided, do students' benefit more from multiple graphical representations than from a single graphical representation. Even though it might seem that Experiment 4 serves to evaluate the claim of my theoretical framework that both connectional sense-making processes and connectional fluency-building processes are *necessary* for students' domain learning to benefit from multiple graphical representations, it is important to note that Experiment 4 did not directly assess whether the proposed learning processes actually take place, but only *infers* from the fact that an instructional intervention designed to support these learning processes is effective that these learning processes need to be supported. It may well be that the sense-making support and fluency-building support used in Experiment 4 actually enhanced other learning processes that are entirely different in nature than the ones proposed by my theoretical framework. Yet, given that prior research on learning with multiple representations has emphasized the importance of supporting students in making sense of the connections between different representations (Bodemer & Faust, 2006; Bodemer et al., 2004; Plötzner et al., 2001; Schwonke et al., 2008; Seufert, 2003; van der Meij & de Jong, 2006) and in becoming fluent in making these connections (Kellman & Garrigan, 2009; Kellman et al., 2008; Kellman et al., 2009; Massey et al., in press), and given that the types of support for connectional sense-making and fluency-building processes were designed in accordance with these literatures, it seems reasonable to say that Experiment 4 provides some (albeit far from conclusive) support for the hypothesis that connectional sense-making processes and connectional fluency-building processes play a role in students' learning with multiple graphical representations, and that support for these learning processes enhances students’ benefit from multiple graphical representations.

Experiment 5 further investigates the interactions between these two hypothesized learning processes. The theoretical framework as described in section 2.2 makes no specific claims as to how the different proposed learning processes interact. Therefore, Experiment 5 can be considered as extending the theoretical framework. Demonstrating that connectional sense-making support should be provided before connectional fluency-building support, Experiment 5 suggests that connectional understanding enhances connectional fluency-building processes, rather than the other way around. However, the same caveats apply as with the previous experiments: to verify this interpretation, a direct assessment of the hypothesized learning processes would be needed.

Taken together, these considerations about my experimental studies illustrate the merit of the theoretical framework for processes involved in learning with multiple graphical representations. The experiments can by no means confirm or disconfirm the theoretical framework – in fact, they were not designed to do so. Yet, in using the theoretical framework as one possible angle from which to interpret the findings from my experiments, I hope to stimulate future research that will directly assess predictions derived from the theoretical framework. In section 2.2.2, I describe how one might derive testable predictions from the theoretical framework, by making additional assumptions about how learning of domain knowledge takes place within the given domain. It would be interesting for future work to specify expected changes in assessment events that directly measure the proposed processes and empirically test the relation between these processes and students’ learning of domain knowledge.

5.2 *Educational technology perspective*

A further contribution of my dissertation is a successful piece of educational software: the Fractions Tutor. The Fractions Tutor is an intelligent tutoring system that uses multiple graphical representations in a research-based way to support the acquisition of robust, conceptual knowledge about fractions. My research on the Fractions Tutor extends the literature on intelligent tutoring systems in various ways described below. The design process of the Fractions Tutor integrates multiple methods from learning sciences, intelligent tutoring systems, and human-computer interaction. In particular, my research illustrates how principle-based integration of these different disciplines can inform the development educational technologies.

5.2.1 *A successful intelligent tutoring system for fractions learning*

The Fractions Tutor is the outcome of a sequence of iterative classroom experiments and lab-based studies. The motivation in developing the Fractions Tutor was to help students acquire robust conceptual and procedural knowledge about fractions, thereby helping them overcome one of the major stumbling blocks in math education (Boyer et al., 2008; Callingham & Watson, 2004; Kaminski, 2002; Person et al., 2004; Moss, 2005). As the relatively poor performance of the majority of 4th-grade students in the recent 2011 national NAEP math assessment demonstrates (see http://nces.ed.gov/nationsreportcard/), there is a need to develop effective instructional tools to help students overcome their difficulties in learning about fractions. Fractions is not only a math topic that is important in its own right, it also provides a crucial foundation for later learning of algebra and other more advanced topics (NMAP, 2008; Siegler et al., 2010).

The Fractions Tutor uses multiple abstract, interactive graphical representations to support students' conceptual learning, thereby taking a different focus than other intelligent tutoring systems, such as ASSISTments (Heffernan et al., 2012), ActiveMath (Goguadze et al., 2008), or Animalwatch (Beal et al., 2010), which have focused on procedural learning, or incorporate non-interactive, concrete, or only a single graphical representation. The decision to use abstract graphical representations is based on cross-iteration studies and corresponds to Goldstone and Son's (2005) concreteness fading approach. The use of interactive graphical representations is motivated by the math education literature on the advantages of virtual manipulatives to support fractions learning (e.g., Durmus & Karakirik, 2006; Moyer et al., 2002; Reimer & Moyer, 2005).

Finally, the use of multiple graphical representations, rather than a single graphical representation, constitutes the heart of the Fractions Tutor. Originally, the decision to include multiple graphical representations was motivated by the finding that most instructional materials employ a variety of graphical representations of fractions. The advantage of using multiple graphical representations was confirmed by my own prior experimental study (see section 4.1 for Experiment 1), and repeatedly throughout subsequent experiments, described above (see section 4.3 for Experiment 3, see section 4.4 for Experiment 4). Based on a sequence of controlled experiments, I investigated how best to implement multiple graphical representations in the Fractions Tutor so that students can take full advantage of the multiplicity of graphical representations.

The outcome of this research is an intelligent tutoring system for fractions that covers a wide range of topics (see section 3.4) and leads to robust conceptual learning gains in real classroom settings (see section 3.5). Given students' difficulties with fractions and the importance of fractions for later math learning, the development of a successful educational software for fractions learning is in and by itself an important contribution. In addition, my approach to integrating learning sciences research with iterative development of educational software based on methods originating in intelligent tutoring systems research and human-computer interaction constitutes a further contribution of my work. I expect that the same approach can benefit the development of other types of educational software as well.

5.2.2 Integrating learning sciences and intelligent tutoring systems research

Throughout my thesis, I have repeatedly emphasized that the Fractions Tutor is both the *outcome* and the *platform* of my research. This emphasis is not incidental; it illustrates an important aspect of my work.

First, my empirical research was guided by theoretically motivated questions about how best to enhance students' benefit from multiple graphical representations. By investigating the effects of instructional support for learning with multiple graphical representations not only on measures of learning outcomes but also on measures of learning processes (e.g., on problem-solving behaviors assessed by tutor log data, or on cognitive processes assessed by interview data and think-aloud protocols), my research provides insights into which learning processes educational technologies need to support through instructional design. Since these learning processes are not specific to fractions, I expect that they play a key role in learning with multiple graphical representations in the numerous other domains that use multiple graphical representations with the same purpose as in fractions: to emphasize complementary conceptual aspects of the domain. Furthermore, my theoretical framework provides a novel perspective that allows to integrate my experimental findings and to deduct specific predictions that can be tested empirically. Let me illustrate how my theoretical framework and the empirical findings from my experimental studies might lead to novel questions and predictions for chemistry learning. Chemistry uses a variety of graphical representations of molecules. Corey-Pauling-Koltun (CPK) representations make the concept of molecule external surface easily accessible, but make it more difficult to perceive the details of the chemical structure. Ball-and-stick figures show the complete chemical structure of a molecule and provide easily recognizable patterns, but they disguise the surface of the molecule. Based on my theoretical framework, I hypothesize that when provided with these different graphical representations of molecules in chemistry instruction, students need to understand how to conceptually interpret each of them, they need to become fluent in using each of them to solve chemistry tasks, and they need to conceptually understand how these different graphical representations relate to one another, and they need to become fluent in establishing these relations. Consequently, students should show higher learning gains in robust chemistry knowledge if these learning processes are supported than when not receiving such support. As this hypothetical example illustrates, one can derive testable predictions from my theoretical framework about which learning processes should be supported so that students' acquisition of robust domain knowledge benefits from multiple graphical representations in domains other than fractions learning.

Thus, by virtue of being theoretically motivated, the investigation of learning sciences questions goes beyond solving a specific implementation problem (such as how best to implement multiple graphical representations within intelligent tutoring systems). Rather, my research provides an empirically evaluated a set of instructional design principles for educational software that uses multiple graphical representations. By investigating the effectiveness of different types of instructional support for learning with multiple graphical representations, my research provides directions for which types of support are effective in helping students take advantage of the different conceptual perspectives that multiple graphical representations provide on complex topics. These principles were evaluated in controlled experiments situated within real classrooms, and can be integrated into other educational technologies and be empirically tested in other domains. Altogether, by using an intelligent tutoring system as a research platform for learning sciences questions, I not only provide insights into learning processes, but also provide practical guidance for the instructional design of multi-representational educational technologies.

Most importantly, the integration of learning sciences research and intelligent tutoring systems research crucially benefits from my use of a multi-methods approach that integrates outcome measures and process-level measures. Throughout my work, I have investigated not only *what works*, but also *how* and *why* it works. Thereby, my research relates instructional design principles to the learning processes they support. For instance, the use of causal path analysis modeling in Experiments 4 and 5 (see sections 4.4.3.2 and 4.5.3.6) helped identify (at least some of) the mechanisms by which sense-making processes and fluency-building processes interact. The use of interview data in Experiment 5 provides insights into the complementary effects of connectional sense-making support and connectional fluency-building support on students' reasoning about domain concepts (see section 4.5.3.4). Understanding which learning processes different types of instructional support enhance allows developers of educational technologies to make an informed decision about which types of instructional support to prioritize. Consider the hypothetical example that a designer of educational software knows that college students tend not to spontaneously make connections between bar charts, box plots, scatter plots, whereas they already have a good representational understanding and representational fluency (having worked each of these representations since high school). Based on my findings, the developer might prioritize on including connectional sense-making support and connectional fluency-building support rather than support for representational understanding and representational fluency.

5.2.3 A principled methodology to resolving design conflicts

Another important aspect of my research is the integration of a particular human-computer interaction perspective, which has substantially contributed to the success of the Fractions Tutor within the context of real educational settings: a principled methodology to resolve design conflicts between stakeholders. Throughout the design process, I was faced with critical design conflicts between competing stakeholder goals. In section 3.3, I describe several options for resolving these conflicts. At the heart of this process is a goal hierarchy that I developed using a bottom-up process that involved a variety of user-centered design methods such as focus groups and affinity diagramming, which allowed to integrate a variety of stakeholders, including students, teachers, and educational psychology experts. Furthermore, I employed a multi-methods approach to empirical research to resolve persisting design conflicts based on parametric experiments and cross-iteration studies. My methodology thereby extends existing instructional design processes by integrating methods from multiple disciplines and addresses shortcomings resulting from their focus on user-centered design alone (Design-based Research Collective, 2003; Jackson et al., 1998; Soloway et al., 1996), learning sciences (Bereiter & Scardamalia, 2003), and cognitive psychology research (Koedinger, 2002; Mayer, 2003; van Merrienboër et al., 2002).

Even though, at times, design decisions are situational, highly contextualized and occur under the pressure of deadlines and therefore are bound to be (to some extent) based on intuition rather than on methods, my approach addresses the common scenario in which developers of educational technologies need to rely on ad-hoc methods to resolve conflicts between conflicting goals of multiple stakeholders. Although I developed and demonstrated this approach within the context of a Cognitive Tutor, a specific type of educational technology that is widely used across 3,000 schools in the United States, I am confident that my approach will generalize to other types of educational technologies. For instance, MIT’s edX system, an open-source learning technology that makes course materials at the college level accessible online, faces unique design challenges due to the learners’ contexts and goals. Users may be students from around the world using the system for exam preparation, or teachers who access the system in order to fulfill their continued education requirement. Conflicts might exist between the users’ goal to relate the learning content to specific contexts, such as for an engineering project (if the user is a college student majoring in engineering), or for a high-school classroom (if the user is a teacher). Addressing these goals is difficult because tailoring the content to these different interest groups would result in having highly specific content that is not at the same time relevant to all interest groups. Yet, MIT has an interest in the edX system being widely used across different groups of users. In applying my approach to combine a goal hierarchy for different types of users with parametric experiments and cross-iteration studies, trade-offs such as the one just described can be explicitly identified and addressed.

The scenario with edX illustrates that the approach I described in this section might serve as a framework to stimulate future research on educational technology development, not only to improve specific technologies, but also to evaluate and further extend the presented approach. Only with a well-researched and principled approach to incorporating multiple (and, as described, often conflicting) stakeholders' goals can we develop educational technologies that are not only effective, but also usable within real contexts and even enjoyable.

5.3 Limitations and future research

Inevitably, as all research, my research is limited in several ways. But at the same time, these limitations point to open questions that will hopefully stimulate further research in the area of learning with multiple graphical representations.

First, my research was conducted in the domain of fractions learning. As mentioned, the way in which fractions instruction employs multiple graphical representations is characteristic of many STEM domains: across many domains, multiple graphical representations are used with the goal to emphasize particular conceptual aspects of the domain that are complementary in the sense that only if students understand all of them, they can be said to have robust domain knowledge. I expect that wherever multiple graphical representations are employed with this purpose, students' benefit from multiple graphical representations will be enhanced if they are employed in accordance with the instructional design principles summarized in section 5.1.1. Based on my theoretical framework, I further hypothesize that students' benefit from multiple graphical representations will be enhanced if they receive instructional support to help them engage in representational sense-making and fluency-building processes as well as in connectional sense-making and fluency-building processes. Yet, this assertion remains to be tested empirically. Future research should investigate whether students' benefit from multiple graphical representations in robust learning in other domains than fractions depends on instructional support for each of these four learning processes. Such research would contribute to further expanding prior research that has focused on learning with multiple representations from different symbol systems (such as text and one additional graphical representation), to learning with multiple representations from the same symbol system (i.e., *multiple graphical* representations). As detailed in section 2.2.2, my theoretical framework allows to deduct hypotheses that can be empirically investigated.

Second, my research is incremental in multiple ways. Not only did the design of instructional support in each experimental study build on the findings in the previous experiments, but also the design of the Fractions Tutor changed between subsequent experiments, based on both the experiment results and on user-based studies conducted within the laboratory. Therefore, the conclusions from each experiment cannot be considered independently from the results of previous experiments. For example, in Experiment 2 (see section 4.2) would interleaving of task types have been more effective than interleaving graphical representations if the reflection prompts that were found to support sense-making of individual graphical representations in Experiment 1 (see section 4.1) had not been integrated in the Fractions Tutor? My research cannot speak to this question. Or, in Experiment 3 (see section 4.3), would interleaving graphical representations have been more effective than blocking graphical representations if task types had not consistently been provided in a fully interleaved fashion across all conditions, based on the findings of Experiment 2 (see section 4.2)? Again, my research does not answer that question. Speaking in terms of my theoretical framework (if one were to adopt my interpretation of the experimental results described in section 5.1.3), it remains an open question whether support for representational fluency-building processes is effective independently of students receiving support for sense-making processes with individual representations. Furthermore, it remains open whether support for learning processes involved in connection making is effective when provided independently of support for learning processes involved in using individual graphical representations. Future research should address these open questions by investigating the effectiveness of each type of instructional support independent of other types of support.

Third, as a consequence of the iterative development of the Fractions Tutor across the experimental studies, the assessments used in each experiment are not directly comparable, as they reflect the changes made to the topics covered by the Fractions Tutor. Inconsistencies between experiments with regard to the results might thus be caused by changes in the structure of the tests. For example, Experiment 4 did not replicate the advantage of multiple graphical representations over a single graphical representation when provided without connection-making support, which was found in Experiment 1 and (to a certain extent) in Experiment 3. Whether this lack of an effect in Experiment 4 might be explained by the fact that the test used in Experiment 4 included both reproduction and transfer items of conceptual knowledge, whereas the test scale on which an advantage of multiple graphical representations was found in Experiment 3 included only transfer items of conceptual knowledge, remains an open question.

Fourth, my research was conducted within a very specific type of educational technology: an intelligent tutoring system. The main strength of intelligent tutoring systems is that they can provide individualized adaptive feedback and hints on demand in real time. Many other educational technologies do not have that capability, such as massive open online courses (MOOCs, like the edX system mentioned earlier). It remains open whether my findings generalize to other technologies that have fewer capabilities to allow for individualized support, or even to allow for the same amount of interactivity with graphical representations. Likewise, it remains open whether my findings generalize to non-technology learning materials, such as paper-based curricula, or paper-and-pencil work sheets.

Finally, although the Fractions Tutor has been shown to lead to robust, substantial learning gains, it could be better. In particular, it does not yet take full advantage of the capability of intelligent tutoring systems to provide adaptive support based on knowledge tracing. Although Cognitive Tutors have the capability to provide adaptive support, I did not make use of this feature because it would have introduced additional (and undesirable) variability in my experimental studies and might have jeopardized the controlled experimental design. An interesting direction for future research would be to investigate the effects of providing instructional support for the hypothesized processes involved in learning with multiple graphical representations based on a knowledge-tracing model. This future version of the Fractions Tutor could provide, for instance, instructional support for fluency-building processes in connection making when students have received mastery in conceptual understanding of the connections. One might envision a domain-independent model that serves as a basis to provide the appropriate type of instructional support for a specific learning process involved in learning with multiple graphical representations when needed. Such a model has the potential to enhance students' benefit from multiple graphical representations in a variety of domains.

5.4 *Summary: An interdisciplinary multi-methods approach*

Although limited in several ways, my research makes important contributions to the fields of learning sciences and educational technologies, as detailed in this section. Furthermore, my research leads to a set of new research questions that will (definitely) stimulate my own future research, and (hopefully) also that of other researchers.

What has substantially influenced these contributions is a particular overarching characteristic of my research: the use of a multi-methods approach. By integrating methods that combine learning outcomes and process-level measures, my research yields insights that motivate novel research questions that can be tested experimentally. To illustrate this point, consider the sequence of experimental studies described in section 4. For instance, the finding in Experiment 3, that students do not spontaneously make connections between graphical representations, unless explicitly prompted to do so (see section 4.3.2) informed the design of connectional sense-making support used in Experiment 4, which includes explicit reflection prompts (see sections 3.4.1 and 3.4.2). Furthermore, the finding in Experiment 4 that connectional sense-making and fluency-building processes interact (see section 4.4.3), together with the result from causal path analysis in Experiment 4 that sense-making support seemed to enhance students' benefit from fluency-building support rather than vice versa (see section 4.4.3.2) lead to two competing hypotheses about learning processes that I contrasted experimentally in Experiment 5. The results from Experiment 5, in turn, lead to the formulation of instructional design principles that I incorporated in the Fractions Tutor and which can be applied to other multi-representational educational technologies as well.

Taken together, these observations illustrate that the combination of learning sciences research and intelligent tutoring systems research yields "more than the sum of their parts". In combining both perspectives, my research benefited from their complementary views: by iteratively moving from theoretically motivated learning sciences questions to implementation questions in the context of intelligent tutoring systems development, both perspectives complemented one another. The result of this integration is (1) a set of empirically validated instructional design principles that can be integrated in a wide range of educational technologies, (2) an empirically motivated theoretical framework for learning with multiple graphical representations, (3) an effective intelligent tutoring system for fractions learning that leads to robust and flexible domain knowledge, and (4) a new methodology that can guide instructional designers in resolving conflicts that inevitably arise in the context of real educational settings.

List of Tables

List of Figures

Acknowledgements

This research would not have been possible without the extraordinary support of my academic advisors, Vincent Aleven and Nikol Rummel, and the immense guidance from my thesis committee members, Richard Scheines, Ken Koedinger, and Brian Junker.

I would also like to thank the Fractions Tutor research group, including Jay Raspat, Michael Ringenberg, Stacie Rohrbach, Zelha Tunc-Pekkan, Laurens Feenstra, Cathy Kramer, Mitch Nathan, Peg Smith, Cassandra Studer, Jenny Olsen, and Dan Belenky. I would also like to thank Zach Pardos for his collaboration on my research. Further, I thank everyone who has given helpful informal comments, including Ryan Baker, Neil Heffernan, Shaaron Ainsworth, Janice Gobert, and David Pritchard.

Further, the Program for Interdisciplinary Educational Research (PIER, lead by David Klahr and Sharon Carver), has fundamentally shaped my research. I would also like to thank others involved in PIER, including Howard Seltman, Jack Mostow, Anna Fisher, all the PIER students, and Audrey Russo.

The Pittsburgh Science of Learning Center (PSLC) has also substantially contributed to my research, and I would like to particularly thank Michael Bett, Gail Kusbit, and Jo Bodnar for their help and support.

I would like to thank everyone who has provided their help within in the Human-Computer Interaction Institute, above all Queenie Kravitz, (again) Jo Bodnar, Lisa Vento, Mark Penney, Katelyn Sittler

Laura Butler, Anthony DeMauro, Adriana Joazeiro, Julika Lomas, Corinne Rockoff, Shawn Snyder, Orly Stampfer, Cassandra Studer, and Angela Wagner deserve special thanks for their help in implementing studies in classrooms.

The PSLC's Datashop and CTAT teams have provided tremendous support for my work, especially Jonathan Sewall, Alida Skogsholm, Brett Leber, Octav Popescu, Cindy Tipper, John Stamper, Kyle Cunningham, Jessica Kalka, Mike Komisin, and Martin van Velsen.

Special thanks goes to interns and research assistants who have helped with this research, including Antonia Becher, Laura Pacilio, Lyubov Zeylikman, Lisa Kwon, Angela McCarthy, Lavender Yi, and Siyan Zhao.

I particularly thank the students and teachers who have participated in my research.

Further, I thank the funding resources for this research, which was provided by the National Science Foundation REESE-21851-1-1121307, the Institute of Education Sciences R305A120734, and by the Pittsburgh Science of Learning Center which is funded by the National Science Foundation, award number SBE-0354420.

Finally, I am very grateful for all the support I have received from my parents, Andrea and Martin Rau, throughout the duration of my PhD.

Literature

Adauto, G., & Klein, J. (2010). *Motion Math: Perceiving fractions through embodied, mobile learning*.

Ainsworth, S. (1999). *Designing effective multi-representational learning environments*. Nottingham: ESRC Centre for Research in Development, Instruction & Training Department of Psychology.

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., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science: A Multidisciplinary Journal, 27*(4), 669-681.

Ainsworth, S. E., Bibby, P. A., & Wood, D. J. (1998). Analysing the Costs and Benefits of Multi-Representational Learning Environments. In M. W. van Someren, P. Reimann, H. P. A. Boshuizen & T. de Jong (Eds.) *Learning with Multiple Representations*. Oxford: Pergamon.

Akpinar, Y., & Hartley, J. R. (1996). Designing Interactive Learning Environments. *Journal of Computer Assisted Learning, 12*(1), 33-46.

Aleven, V. (2010). Rule-based cognitive modeling for intelligent tutoring systems. In *Advances in Intelligent Tutoring Systems* (pp. 33-62). Berlin / Heidelberg: Springer.

Aleven, V., & Koedinger, K. R. (2000). The need for tutor dialog to support self-explanation. In C. P. Rose & R. Freedman (Eds.), Building Dialogue Systems for Tutorial Applications. *Papers from the 2000 AAAI Fall Symposium* (pp. 65-73). Menlo Park, CA: AAAI Press.

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. (2006). The Cognitive Tutor Authoring Tools (CTAT): Preliminary evaluation of efficiency gains. In *Intelligent Tutoring Systems* (pp. 61-70). Berlin / Heidelberg: Springer.

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009a). Example-tracing tutors: A new paradigm for intelligent tutoring systems. *International Journal of Artificial Intelligence in Education, 19*(2), 105-154.

Aleven, V., McLaren, B. M., & Sewall, J. (2009b). Scaling up programming by demonstration for intelligent tutoring systems development: An open-access website for middle-school mathematics learning. *IEEE Transactions on Learning Technologies, 2*(2), 64-78.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Anderson, J. R. (2002). *ACT: A simple theory of complex cognition.* Cambridge, MA: MIT Press.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, *4*(2), 167-207.

Anderson, J. R., & Gluck, K. (2001). What role do cognitive architectures play in intelligent tutoring systems. *Cognition & Instruction: Twenty-five years of progress*, 227-262.

Arcavi, A. (2003). The Role of Visual Representations in the Learning of Mathematics. *Educational Studies in Mathematics, 52*(3), 215-241.

Arroyo, I., Royer, J. M., & Woolf, B. P. (2011). Using an intelligent tutor and math fluency training to improve math performance. *International Journal of Artificial Intelligence in Education, 21*(1), 135-152.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181-214.

Atkinson, R. K., & Renkl, A. (2007). Interactive example-based learning environments: Using interactive elements to encourage effective processing of worked examples. *Educational Psychology Review*, *19*(3), 375-386.

Baetge, I., & Seufert, T. (2010). Effects of support for coherence formation in computer science education. Paper presented at the EARLI SIG 6/7.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of Foreign-Language Vocabulary and the Spacing Effect. *Psychological Science, 4*(5), 316-321.

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2002). The Resilience of Overgeneralization of Knowledge about Data Representations. Paper presented at the American Educational Research Association Conference.

Baker, R. S. J. d., Corbett, A. T., & Koedinger, K. R. (2007). The difficulty factors approach to the design of lessons in intelligent tutor curricula. *International Journal of Artificial Intelligence in Education, 17*(4), 341-369.

Ball, D. (1990). *Halves, Pieces, and Twoths: Constructing Representational Contexts in Teaching Fractions*. East Lansing: Michigan State University, National Center for Research on Teacher Education.

Bannert, M. (2002). Managing cognitive load-recent trends in cognitive load theory. *Learning and Instruction*, *12*(1), 139-146.

Bassok, M. (1996). Using content to interpret structure: Effects on analogical transfer. *Current Directions in Psychological Science*, *5*(2), 54-58.

Battig, W. F. (1972). Intratask interference as a source of facilitation in transfer and retention. In R. F. Thompson & J. F. Vos (Eds.), *Topics in learning and performance* (pp. 131-159). New York: Academic Press.

Beal, C. R., Arroyo, I. M., Cohen, P. R., Woolf, B. P., & Beal, C. R. (2010). Evaluation of AnimalWatch: An intelligent tutoring system for arithmetic and fractions. *Journal of Interactive Online Learning*, *9*(1), 64-77.

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.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55-68.

Bennett, J. M. (2004). *Holt Middle School Math.* Holt, Rinehart & Winston.

Bereiter, C., & Scardamalia, M. (2003). Learning to work creatively with knowledge. In E. de Corte (Ed.), *Powerful learning environment* (pp. 55-68). Oxford: Pergamon.

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, 37*(4), 345-363.

Berthold, K., Faulhaber, L., Guevara, T., & Renkl, A. (2010). Training focused processing of explanations: How can transfer across domains be fostered? Paper presented at the EARLI SIG 6/7.

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

Beyer, H., & Holtzblatt, K. (Eds.). (1998). *Contextual Design: Defining Customer-Centered Systems*. San Francisco, CA: Morgan Kaufmann Publishers.

Bieda, K. N., & Nathan, M. J. (2009). Representational disfluency in algebra: evidence from student gestures and speech. *ZDM, 41*(5), 637-650.

Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(3), 792-810.

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

Bodemer, D., Plötzner, R., Bruchmüller, K., & Häcker, S. (2005). Supporting learning with interactive multimedia through active integration of representations. *Instructional Science, 33*(1), 73-95.

Bodemer, D., Plötzner, 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.

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental psychology, 44*(5), 1478-1490.

Brenner, M. E., Mayer, R. E., Moseley, B., Brar, T., Duran, R., Reed, B. S. et al. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal, 34*(4), 663-689.

Brinker, L. (1997). *Using Structured Representations To Solve Fraction Problems: A Discussion of Seven Students' Strategies*.

Brünken, R., Seufert, T., & Zander, S. (2005). Förderung der Kohärenzbildung beim Lernen mit multiplen Repräsentationen. *Zeitschrift für Pädagogische Psychologie, 19*(1), 61-75.

Butcher, K., & Aleven, V. (2007). Integrating visual and verbal knowledge during classroom learning with computer tutors. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 137-142). New York, NY: Lawrence Erlbaum.

Butcher, K., & Aleven, V. (2008). Diagram interaction during intelligent tutoring in geometry: Support for knowledge retention and deep transfer. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*. New York, NY: Lawrence Erlbaum.

Butcher, K. R. (2006). Learning From Text With Diagrams: Promoting Mental Model Development and Inference Generation. *Journal of Educational Psychology, 98*(1), 182-197.

Byerly, G. (2007). Look in their eyes - Eye tracking, Usability, and Children. *School Library Media Activities Monthly, 23*(8), 30-32.

Caldwell, J. H. (1995). Communicating about fractions with pattern blocks. *Teaching children mathematics, 1*(9), 156-161.

Callingham, R., & Watson, J. (2004). A Developmental Scale of Mental Computation with Part-Whole Numbers. *Mathematics Education Research Journal, 16*(2), 69-86.

Caney, A., & Watson, J. M. (2003). Mental Computation Strategies for Part-Whole Numbers. Paper presented at the International Educational Research Conference, Auckland.

Carmichael, A., Larson, A., Gire, E., Loschky, L., & Rebello, N. S. (2010). How Does Visual Attention Differ Between Experts and Novices on Physics Problems? In *AIP Conference Proceedings* (pp. 93-96).

Carpenter, T. P. (1971). *The Performance of First Grade Students on a Nonstandard Set of Measurement Tasks*.

Catrambone, R., & Holyoak, K. J. (1998). Learning and subgoals and methods forsolving probability problems. *Memory & Cognition, 18*, 593-603.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354-380.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction, 8*(4), 293-332.

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.

Cheng, P. (1999). Unlocking conceptual learning in mathematics and science with effective representational systems. *Computers & Education, 33*, 109-130.

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

Chiu, J. L., & Linn, M. C. (2012). The role of self-monitoring in learning chemistry with dynamic visualizations. In *Metacognition in Science Education* (pp. 133-163). Springer Netherlands.

Clark, R. C., & Mayer, R. E. (2003). Applying the modality principle. In *E-Learning And The Science Of Instruction: Proven Guidelines For Consumers And Designers Of Multimedia Learning* (pp. 83-95). San Francisco, CA: Pfeiffer.

Clark, R. E., Feldon, D. F., van Merrienboër, J. J. G., Yates, K. A., & Early, S. (2007). Cognitive Task Analysis. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cobb, P. (1995). Cultural tools and mathematical learning: A case study. *Journal for Research in Mathematics Education, 26*(4), 362-385.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Conati, C., Merten, C., Muldner, K., & Ternes, D. E. (2005). Exploring eye tracking to increase bandwidth in user modeling. In *User Modeling* (pp. 357-366). Berlin / Heidelberg: Springer

Confrey, J., & Maloney, A. (2010). The construction, refinement, and early validation of the equipartitioning learning trajectory. In *Proceedings of the 9th International Conference of the Learning Sciences Vol. 1* (pp. 968-973). International Society of the Learning Sciences.

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.

Corbett, A., & Anderson, J. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In *Proceedings of the SIG CHI Conference on Human Factors in Computing Systems* (pp. 245-252). ACM Press.

Corbett, A., Kauffman, L., Maclaren, B., Wagner, A., & Jones, E. (2010). A Cognitive Tutor for genetics problem solving: Learning gains and student modeling. *Journal of Educational Computing Research, 42*(2), 219-239.

Corbett, A., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. In M. Helander, T. K. Landauer & P. Prabhu (Eds.), *Handbook of human-computer interaction, second, completely revised edition* (pp. 849-874). New York: Elsevier Science Publishers.

Corbett, A., & Trask, H. (2000). Instructional interventions in computer-based tutoring: Differential impact on learning time and accuracy. In *In Proceedings of the SIG CHI Conference on Human Factors in Computing Systems* (pp. 94-104). ACM Press.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*, 253-278.

Corbett, A. T., Koedinger, K., & Hadley, W. S. (2001). Cognitive tutors: From the research classroom to all classrooms. In P. S. Goodman (Ed.), *Technology enhanced learning: Opportunities for change* (pp. 235-263). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

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

Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction, 9*(4), 343-363.

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

Cramer, K., Behr, M., Post, R., & Lesh, R. (1997a). *Rational Number Project: Fraction Lessons for the Middle Grades: Level 1*. Dubuque, IA: Kendall/Hunt Publishing.

Cramer, K., Behr, M., Post, R., & Lesh, R. (1997b). *Rational Number Project: Fraction Lessons for the Middle Grades: Level 2*. Dubuque, IA: Kendall/Hunt Publishing.

Cramer, K., & Henry, A. (2002). Using Manipulative Models to Build Number Sense for Addition of Fractions. In B. Litwiller & G. Bright (Eds.), *Making Sense of Fractions, Ratios, and Proportions*. Reston, VA: National Council of Teachers of Mathematics.

Cramer, K., & Wyberg, T. (2007). When getting the right answer is not always enough: Connecting how students order fractions and estimate sums and differences. In *The Learning of Mathematics: Sixty-ninth Yearbook* (pp. 205-220). Reston, VA: The National Council of Teachers of Mathematics.

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), 490-496.

Cramer, K., Wyberg, T., & Leavitt, S. (2009). Rational Number Project. Fraction Operations & Initial Decimal Ideas. Dubuque, IA: Kendall/Hunt Publishing.

Cramer, K. A., Post, T. R., & delMas, R. C. (2002). Initial Fraction Learning by Fourth- and Fifth-Grade Students: A Comparison of the Effects of Using Commercial Curricula with the Effects of Using the Rational Number Project Curriculum. *Journal for Research in Mathematics Education, 33*(2), 111-144.

Crawford, C., & Brown, E. (2003). Integrating internet-based mathematical manipulatives within a learning environment. *Journal of Computers in Mathematics and Science Teaching, 22*(2), 169-180.

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Cognitive activities in complex science text and diagrams. *Contemporary Educational Psychology, 35*(1), 59-74.

de Croock, M. B. M., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). High versus low contextual interference in simulation-based training of troubleshooting skills: Effects on transfer performance and invested mental effort. *Computers in Human Behavior, 14*(2), 249-267.

de Jong, T., Ainsworth, S. E., Dobson, M., Van der Meij, J., Levonen, J., Reimann, P. et al. (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* (pp. 9-41). Amsterdam: Elsevier Science.

de Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction, 20*(2), 111-122.

DeBell, M., Chapman, C., Kleiner, A., Lewis, L., Hudson, L., & Shafer, L. (2003). *Computer and Internet Use.* U.S. Department of Education.

Design-based Research Collective. (2003). An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Duchowski, A. (2007). *Eye Tracking Techniques Eye Tracking Methodology* (pp. 51-59). London: Springer.

Durmus, S., & Karakirik, E. (2006). Virtual manipulatives in mathematics education: A theoretical framework. *The turkish online journal of educational technology, 5*(1), 117-123

Eilam, B. (2013). Possible Constraints of Visualization in Biology: Challenges in Learning with Multiple Representations. In D. F. Treagust & C.-Y. Tsui (Eds.) *Multiple Representations in Biological Education* (pp. 55-73). Springer Netherlands.

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

Eitel, A., Scheiter, K., & Schüler, A. (2013). How Inspecting a Picture Affects Processing of Text Applied Cognitive Psychology. *Applied Cognitive Psychology*. doi: 10.1002/acp.2922

Even, R. (1998). Factors involved in linking representations of functions. *The Journal of Mathamtical Behavior, 17*(1), 105-121.

Fahle, M., & Poggio, T. (2002). *Perceptual Learning*. MIT Press.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Educational Technology & Society, 32*, 124-158.

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York, NY: Springer-Verlag Publishing.

Fuson, K. C., & Abrahamson, D. (2005). Understanding Ratio and Proportion as an Example of the Apprehending Zone and Conceptual-Phase Problem-Solving Models. In J. I. D. Campbell (Ed.) *Handbook of Mathematical Cognition*. New York, NY: Psychology Press.

Gagatsis, A., & Elia, I. (2004). The effects of different modes of representation on mathematical problem solving. In *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (pp. 447-454).

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review, 23*(4), 523-552.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*(1), 45-56.

Gerjets, P., Schwonke, R., & Catambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction*, *16*, 104-121.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Psychology Press.

Gilbert, J. K. (2008). Visualization: An emergent field of practice and inquiry in science education. In J. K. Gilbert, M. Reiner & M. B. Nakhleh (Eds.) *Visualization: Theory and practice in science education*. Springer.

Gire, E., Carmichael, A., Chini, J. J., Rouinfar, A., Rebello, S., Smith, G. et al. (2010). The effects of physical and virtual manipulatives on students' conceptual learning about pulleys. In *Proceedings of the 9th International Conference of the Learning Sciences Vol. 1* (pp. 937-943). International Society of the Learning Sciences.

Gobert, J. D., O'Dwyer, L., Horwitz, P., Buckley, B. C., Levy, S. T., & Wilensky, U. (2011). Examining the Relationship Between Students' Understanding of the Nature of Models and Conceptual Learning in Biology, Physics, and Chemistry. *International Journal of Science Education, 33*(5), 653-684.

Goguadze, G., Melis, E., & DFKI. (2008). Feedback in ActiveMath Exercises. In *Proceedings of the International Conference on Mathematics Educatio*n.

Goldman, S. (2003). Learning in complex domains: When and why do multiple representations help? *Learning and Instruction, 13*(2), 239-244.

Goldstone, R. L., & Son, J. Y. (2005). The Transfer of Scientific Principles Using Concrete and Idealized Simulations. *Journal of the Learning Sciences, 14*(1), 69-110.

Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition, 31*(2), 169-180.

Grady, S. M. (1998). *Virtual reality: Computers mimic the physical world*. New York: University Press.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17*(6), 612-634.

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.

Hake, S. (2004). *Saxon Math 5/6*: Saxon Pub.

Halpern, D. F., Aronson, J., Reimer, N., Simkins, S., Star, J. R., & Wentzel, K. (2007). *Encouraging Girls in Math and Science*. Washington, DC: Institute of Educational Sciences, U.S. Department of Education.

Harden, R. M., & Stamper, N. (1999). What is a spiral curriculum? *Medical Teacher, 21*(2), 141-143.

Hebert, E. P., Landin, D., & Solmon, M. A. (1996). Practice schedule effects on the performance and learning of low- and high-skilled students: An applied study. *Research Quarterly for Exercise and Sport, 67*(1), 52-58.

Hecht, S. A., Vagi, K. J., Torgesen, J. K., Berch, D. B., & Mazzocco, M. l. M. M. (2007). Fraction skills and proportional reasoning. In *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 121-132). Baltimore, MD: Paul H Brookes Publishing.

Heffernan, N. T., Heffernan, C., Deceoteu, M., & Militello, M. (2012). Effective and meaningful use of educational technology: three cases from the classroom. In C. Dede & J. Richards (Eds.) *Digital Teaching Platforms* (pp. 88–102). New York: Teacher's College Press.

Heim, M. (2000). *Virtual realism*. New York: Oxford University Press.

Helsdingen, A., van Gog, T., & van Merriënboer, J. J. G. (2011). The effects of practice schedule and critical thinking prompts on learning and transfer of a complex judgement task. *Journal of Educational Psychology, 103*(2), 383-398.

Heo, J.-J., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 37-53.

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction, 18*(1), 54-65.

Hinze, S. R., Rapp, D. N., Williamson, V. M., Shultz, M. J., Deslongchamps, G., & Williamson, K. C. (2013). Beyond ball-and-stick: Students' processing of novel STEM visualizations. *Learning and Instruction, 26*, 12-21.

Holsanova, J., & Holmberg, N. H., K. (2009). Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology, 23*, 1215–1226.

Hwang, W.-Y., Su, J.-H., Huang, Y.-M., & Dong, J.-J. (2009). A study of multi-representation of geometry problem solving with virtual manipulatives and whiteboard system. *Educational Technology & Society, 12*(3), 229-247.

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*(2), 172-176.

Hyönä, J., & Lorch, R. F. (2004). Effects of topic headings on text processing: Evidence from adult readers' eye fixation patterns. *Learning and Instruction, 14*(2), 131-152.

Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye movement measures to study global text processing. In *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313-334).

Immink, M. A., & Wright, D. L. (1998). Contextual interference: A response planning account. *The Quarterly Journal of Experimental Psychology Section A, 51*(4), 735-754.

Irwin, D. E. (2004). Fixation Location and Fixation Duration as Indices of Cognitive Processing. In Henderson, J. M., & Ferreira, F. (Eds.) *The Ingegration of Language, Vision, and Action: Eye Movements and the Visual World* (pp. 105-133). New York, NY: Psychology Press.

Jackson, S., Krajcik, J. S., & Soloway, E. (1998). The design of guided learner-adaptable scaffolding in interactive learning environments. In *Proceedings of the SIG CHI Conference on Human Factors in Computing Systems* (pp. 187-194). ACM Press.

Jakobsdottir, S., Krey, C. L., & Sales, G. C. (1994). Computer graphics: Preferences by gender in grades 2, 4, and 6. *Journal of Educational Research, 88*(2), 91-100.

Jarodzka, H., van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a Model's eye movements fosters learning. 25, 62-70.

Jenkinson, J. (2009). Measuring the effectiveness of educational technology: What are we attempting to measure? *Journal of e-Learning, 7*(3), 273-280.

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. Journal of Experimental Psychology: Applied, 18(2), 178-191.

Johnson, K. R., & Layng, T. V. J. (1996). On terms and procedures: Fluency. *The Behavior Analyst, 19*(2), 281-288.

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.

Kamii, C., & Clark, F. B. (1995). Equivalent Fractions: Their Difficulty and Educational Implications. *Journal of Mathematical Behavior, 14*(4), 365-378.

Kaminski, E. (2002). Promoting Mathematical Understanding: Number Sense in Action. *Mathematics Education Research Journal, 14*(2), 133-149.

Kapur, M., & Rummel, N. (2009). The assistance dilemma in CSCL. In A. Dimitracopoulou, C. O'Malley, D. D. Suthers & P. Reimann (Eds.), *Proceedings of the Computer-Supported Collaborative Learning Conference Vol. 2* (pp. 37-39). International Society of the Learning Sciences.

Kaput, J. J. (1989). Linking Representations in the Symbol Systems of Algebra. In S. Wagner (Ed.), *Research issues in the learning and teaching of algebra* (pp. 167-195). Reston, VA: National Council of Teachers of Mathematics.

Kellman, P. J., & Garrigan, P. B. (2009). Perceptual learning and human expertise. *Physics of Life Reviews*, *6*(2), 53-84.

Kellman, P. J., Massey, C. M., Roth, Z., Burke, T., Zucker, J., Saw, A. et al. (2008). Perceptual learning and the technology of expertise: Studies in fraction learning and algebra. *Pragmatics & Cognition, 16*(2), 356-405.

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, 2*, 285-305.

Kent, L. B., Arnosky, J., & McMonagle, J. (2002). Using representational contexts to support multiplicative reasoning. In B. Litwiller & G. Bright (Eds.) *Making sense of fractions, ratios, and propoertions* (pp. 145-152). Reston, VA: National Council of Teachers of Mathematics.

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.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Kim, D., Kim, D. J., & Whang, W. H. (2013). Cognitive Synergy in Multimedia Learning. *International Education Studies, 6*(4), 76-85.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*(1), 1-10.

Koedinger, K. R. (2002). Toward evidence for instructional design principles: Examples from Cognitive Tutor Math 6. In *Proceedigns of the 24th Annual Meeting of the North American Chapter of the International Group of the Psychology of Mathematics Education* (pp. 21-49). Athens, GA.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review, 19*(3), 239-264.

Koedinger, K. R., Baker, R., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC Data-Shop. In C. Romero (Ed.), *Handbook of educational data mining* (pp. 10-12). Boca Raton, FL: CRC Press.

Koedinger, K. R., & Corbett, A. (2006). *Cognitive Tutors: Technology Bringing Learning Sciences to the Classroom*. New York, NY: Cambridge University Press.

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.

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences, 13*(2), 129-164.

Koedinger, K. R., Pavlik, P., McLaren, B. M., & Aleven, V. (2008). Is it Better to Give than to Receive? The Assistance Dilemma as a Fundamental Unsolved Problem in the Cognitive Science of Learning and Instruction. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2155-2160). Austin, TX: Cognitive Science Society.

Kong, S. C. (2008). The Development of a Cognitive Tool for Teaching and Learning Fractions in the Mathematics Classroom: A Design-Based Study. *Computers & Education, 51*(2), 886-899.

Kong, S. C., & Kwog, L. F. (2003). A graphical partitioning model for learning common fraction: designing affordances on a web-supported learning environment. *Computers and Education, 40*, 137-155.

Kong, S. C., Lam, S. Y., & Kwog, L. F. (2005). A cognitive tool in handheld devices for collaborative learning: Comprehending procedural knowledge of the addition of common fractions. In *Computer Support for Collaborative Learning,* Taipei, Taiwan.

Kopp, V., Stark, R., & Fischer, M. (2008). Fostering diagnostic knowledge through computer-supportede, case-based worked examples: effects of erroneous examples and feedback. *Medical Education, 42*, 823-329.

Kordaki, M. (2010). A drawing and multi-representational computer environment for beginners' learning of programming using C: Design and pilot formative evaluation. *Computers & Education, 54,*69-87.

Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *Journal of the Learning Sciences, 9*(2), 105-143.

Kozma, R., & Russell, J. (2005). Students becoming chemists: Developing representational competence Visualization in science education (pp. 121-145). Springer Netherlands.

Kuehl, T., Scheiter, K., & Gerjets, P. (2010). Can text compensate differences in types of graphics? Paper presented at the EARLI SIG 2.

Kyun, S. A., & Lee, H. (2009). The effects of worked examples in computer-based instruction: Focus on the presentation format of worked examples and prior knowledge of learners. *Asia Pacific Education Review, 10*(4), 495-503.

Lamberty, K. K., & Kolodner, J. L. (2002). Exploring digital quilt design using manipulatives as a learning tool. In Proceedings of the International Conference of the Learning Sciences Vol. 2 (pp. 552-553). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Lappan, G., Fey, J. T., & Fitzgerald, W. M. (1998). *Bits & Pieces: Understanding Rational Numbers*. Dale Seymour Publications.

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.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(4), 730-746.

Lee, T. D., & Magill, R. A. (1985). Can forgetting facilitate skill acquisition? In D. Goodman, R. B. Wilberg & I. M. Franks (Eds.) *Differing perspectives in motor learning, memory, and control*. Amsterdam: Elsevier.

Leikin, R., Leikin, M., Waisman, I., & Shaul, S. (2013). Effect of the presence of external representations on accurcacy and reaction time in solving mathematical double-choice problems by students of different levels of instruction. *International Journal of Science and Mathematics Education*.

Lesh, R., Post, T., & Behr, M. (1988). Proportional Reasoning. In J. Hiebert & M. Behr (Eds.), *Number Concepts and Operations in the Middle Grades* (pp. 93-118). Reston, VA: National Council of Teachers of Mathematics.

Lewalter, D. (2003). Cognitive Strategies for Learning from Static and Dynamic Visuals. *Learning and Instruction, 13*(2), 177-189.

Lewis, D., & Barron, A. (2009). Animated demonstrations: Evidence of improved performance efficiency and the worked example effect. *Human Centered Design, Lecture Notes in Computer Science, 5619*, 247–255.

Li, Y. H., & Wright, D. L. (2000). An assessment of the attention demands during random- and blocked-practice schedules. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 53*(2), 591-606.

Löhner, S., van Joolingen, W. R., & Savelsbergh, E. R. (2003). The Effect of External Representation on Constructing Computer Models of Complex Phenomena. *Instructional Science, 31*(6), 395-418.

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

Mack, N. K. (1993). Learning rational numbers with understanding: The case of informal knowledge. In T. P. Carpenter, E. Fennema & T. A. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 85-105). Hillsdale, NJ: Lawrence Erlbaum Associates.

Magner, U., Schwonke, R., Renkl, A., Aleven, V., & Popescu, O. (2010). Seductive Illustrations: Double-Edged Effects? Paper presented at the EARLI SIG 6/7.

Martin, T., Svihla, V., & Smith, C. P. (2012). The role of physical action in fraction learning. *Education and Human Development, 5*(1), 1-17.

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.

Mason, L., Pluchino, P., & Tornatora, M. C. (2013). Effects of Picture Labeling on Science Text Processing and Learning: Evidence From Eye Movements. *Reading Research Quarterly, 48*(2), 199-214.

Massey, C., Kellman, P. J., Roth, Z., & Burke, T. (in press). Perceptual learning and adaptive learning technology: developing new approaches to mathematics learning in the classroom. In N. L. Stein & S. W. Raudenbush (Eds.), *Developmental and learning sciences go to school*. New York: Taylor and Francis.

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction, 13*(2), 125-139.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed), *The Cambridge handbook of multimedia learning* (pp. 31-48). Cambridge University Press.

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20*(2), 167-171.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*(4), 715-726.

Mayer, R. E., & Moreno, R. (1998). A cognitive theory of multimedia learning: Implications for design principles. In *Proceedings of the SIG CHI Conference on Human Factors in Computing Systems.* ACM Press.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

McDaniel, M. A., & Butler, A. C. (2010). contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork*. New York: Psychology Press.

McLaren, B. M., Lim, S.-J., & Koedinger, K. R. (2008). When and How Often Should Worked Examples be Given to Students? New Results and a Summary of the Current State of Research. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2176-2181). Austin, TX: Cognitive Science Society.

Meagher, M. (2002). *Teaching Fractions: New Methods, New Resources*.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(6), 1423-1442.

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. Cognitive Psychology, 41(3), 211-253.

Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (1997). Meaningful learning in science: The human constructivist perspective. In G. D. Phye (Ed.), *Handbook of academic learning: Construction of knowledge*. (pp. 405-447). San Diego, CA: Academic Press.

Miura, I. T., & Yamagishi, J. M. (2002). The development of rational number sense. In B. Litwiller & G. Bright (Eds.), *Making sense of fractions, ratios, and proportions*. Reston, VA: National Council of Teachers of Mathematics.

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.

Moyer, P., Bolyard, J., & Spikell, M. A. (2002). What are virtual manipulatives? *Teaching children mathematics, 8*, 372-377.

Murphy, K. (2001). The bayes net toolbox for matlab. *Computing science and statistics, 33*(2), 1024-1034.

Nathan, M. J., Walkington, C. A., Srisurichan, R., & Alibali, M. W. (2011). Modal engagements in precollege engineering: tracking math and science concepts across symbols, sketches, software, silicone and wood. In *Engineering in PreCollege Settings: Research into Practice*. American Society for Engineering Education.

NCTM. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

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

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

NETP. (2010). Tra*nsforming American education: Learning powered by technology*. U.S. Department of Education, Office of Educational Technology.

Ni, Y. (2001). Semantic domains of rational numbers and the acquisition of fraction equivalence. *Contemporary Educational Psychology, 26*(3), 400-417.

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.

Nievelstein, F., van Gog, T., van Dijck, G., & Boshuizen, H. P. A. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology, 38*, 118-125.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Nistal, A., Van Dooren, W., Clarebout, G., Elen, J., & Verschaffel, L. (2009). Conceptualising, investigating and stimulating representational flexibility in mathematical problem solving and learning: a critical review. *ZDM Mathematics Education, 41*, 627-636.

Nistal, A., Van Dooren, W., Clarebout, G., Elen, J., & Verschaffel, L. (2010). Representational flexibility in linear-function problems. In L. Verschaffel, E. De Corte, T. de Jong, & J. Elen (Eds.), *Use of Representations in Reasoning and Problem Solving: Analysis and Improvement*. Oxon: Routledge.

Nistal, A. A., Van Dooren, W., Clarebout, G., Elen, J., & Verschaffel, L. (2010). Representational flexibility in linear-function problems: A choice/no-choice study. In L. Verschaffel, E. De Corte, Elen & T. D. Jong (Eds.), *Representational flexibility in function problems* (pp. 74-93). Oxon: Routledge.

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

Nokes, T. J., & VanLehn, K. (2008). Bridging principles and examples through analogy and explanation. In *Proceedings of the 8th international conference on International Conference for the Learning Cciences Vol. 3* (pp. 100-102). International Society of the Learning Sciences.

Noss, R. R., Healy, L., & Hoyles, C. (1997). The construction of mathematical meanings: Connecting the visual with the symbolic. *Educational Studies in Mathematics, 33*, 203-233.

Nunes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics*. New York, NY: Cambridge University Press.

Ogan, A., Aleven, V., & Jones, C. (2008). Pause, predict, and ponder: Use of narrative videos to improve cultural discussion and learning. In *Proceedings of the SIG CHI Conference on Human Factors in Computing Systems.* ACM Press.

Ohlsson. (1991). Knowledge requirements for teaching: The case of fractions. In P. Goodyear (Ed.), *Teaching knowledge and intelligent tutoring* (pp. 25-59). Norwood, NJ: Ablex.

Oliver, K., Osborne, J., & Brady, K. (2009). What are secondary students' expectations for teachers in virtual school environments? *Distance Education, 30*(1), 23-45.

Ollis, S., Button, C., & Fairweather, M. (2005). The influence of professional expertise and task complexity upon the potency of the contextual interference effect. *Acta Psychologica, 18*(3), 229-244 .

Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding afffects multimedia learning. *Computers & Education, 53*(2), 445–453.

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.

Paas, F., & Van Gog, T. (2006). Optimixing worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction, 16*, 87-91.

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122-133.

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors, 35*(4), 737-743.

Pagni, D. (2004). Fractions and Decimals. *Australian Mathematics Teacher, 60*(4), 28-30.

Paik, J. H. (2005). Fraction concepts: A complex system of mappings. *ProQuest Information & Learning.*

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.

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

Pardos, Z., & Heffernan, N. T. (2010). Navigating the parameter space of Bayesian Knowledge Tracing models: Visualizations of the convergence ofthe Expectation Maximization algorithm. In S. Baker, A. Merceron & P. I. Pavlik (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 161-170).

Pardos, Z. A., Dailey, M. D., & Heffernan, N. T. (2010). Learning what works in ITS from non-traditional randomized controlled trial data. In V. Aleven, J. Kay & J. Mostow (Eds.), *Proceedings of the 10th International Conference on ITS2010* Vol. 6094 (pp. 41-50). Berlin / Heidelberg: Springer.

Pardos, Z. A., & Heffernan, N. T. (2011). KT-IDEM: Introducing Item Difficulty to the Knowledge Tracing Model. In *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 243-254). Berlin / Heidelberg: Springer.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M. et al. (2007). *Organizing Instruction and Study to Improve Student Learning IES Practice Guide*. National Center for Education Research.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. P. (2007). Enhancing Learning and Retarding Forgetting: Choices and Consequences. *Online Submission Description, 19*.

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

Person, A. C., Berenson, S. B., & Greenspon, P. J. (2004). The Role of Number in Proportional Reasoning: A Prospective Teacher's Understanding. In *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education Vol. 4* (pp. 17-24).

Pett, D., & Wilson, T. (1996). Color research and its application to the design of instructional materials. *Educational Technology Research and Development, 44*(3), 19-35.

Pirolli, P. L., & Anderson, J. R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *39*(2), 240-272.

Pitta-Pantazi, D., Gray, E. M., & Christou, C. (2004). Elementary School Students' Mental Representation of Fractions. In *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education Vol. 4* (pp. 17-24).

Plötzner, R., Bodemer, D., & Feuerlein, I. (2001). Facilitating the mental integration of multiple sources of information in multimedia learning environments. In *Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications Vol. 2001* (pp. 1501-1506).

Plötzner, R., Bodemer, D., & Neudert, S. (2008). Successful and less successful use of dynamic visualizations in instructional texts. In R. K. Lowe & W. Schnotz (Eds.), *Learning with Animation. Research implications for design* (pp. 71-91). Cambridge University Press.

Post, T. R., Behr, M. J., & Lesh, R. (1982). Interpretations of Rational Number Concepts. In L. Silvey & J. R. Smart (Eds.), *Mathematics for the Middle Grades (5-9)* (pp. 59-72). Reston, VA: National Council of Teachers of Mathematics.

Preece, J. (1993). Graphs are not straightforward. In T. R. G. Green, S. J. Payne & G. C. van der Veer (Eds.), *The psychology of computer use* (pp. 41-56). London, UK: Academic Press.

Proctor, R. M. J., Baturo, A. R., & Cooper, T. J. (2002). Integrating concrete and virtual materials in an elementary mathematics classroom: A case study of success with fractions. In *Proceedings of the Seventh World Conference on Computers in Education* (pp. 87-92). Australian Computer Society, Inc.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*(1), 144-161.

Rasch, T., & Schnotz, W. (2009). Interactive and non-interactive pictures in multimedia learning environments: Effects on learning outcomes and learning efficiency. *Learning and Instruction*, *19*(5), 411-422.

Rathmell, E. C., & Leutzinger, L. P. (1991). Number representations and relationships. *The Arithmetic Teacher, 38*(7), 20-23.

Rau, M. A. (2008). *Flexible knowledge of fractions with multiple graphical representations in intelligent tutoring systems*. Albert-Ludwigs-Universität.

Rau, M. A., Aleven, V., & Rummel, N. (2009). Intelligent tutoring systems with multiple representations and self-explanation prompts support learning of fractions. In V.

Dimitrova, R. Mizoguchi & B. du Boulay (Eds.), *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling* (pp. 441-448). Amsterdam: IOS Press.

Rau, M. A., Aleven, V., & Rummel, N. (2010). Blocked versus Interleaved Practice with Multiple Representations in an Intelligent Tutoring System for Fractions. In V. Aleven, J. Kay & J. Mostow (Eds.), *Intelligent Tutoring Systems Vol. 6094* (pp. 413-422). Berlin / Heidelberg: Springer

Rau, M. A., Aleven, V., & Rummel, N. (2013a). Complementary effects of sense-making and fluency-building support for connection making: A matter of sequence? To appear in *Proceedings of AIED 2013*.

Rau, M. A., Aleven, V., & Rummel, N. (2013b). Interleaved practice in multi-dimensional learning tasks: which dimension should we interleave? *Learning and Instruction, 23*, 98-114.

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. Cerri, W. Clancey, G. Papadourakis & K. Panourgia (Eds.), *Intelligent Tutoring Systems Vol. 7315* (pp. 174-184). Berlin / Heidelberg: 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. To appear in *Proceedings of the 2013 ACM Conference on Human Factors in Computing Systems*.

Rau, M. A., & Pardos, Z. A. (2012). Investigating Practice Schedules of Multiple Fraction Representations Using Knowledge Tracing Based Learning Analysis Techniques. In K. Yacef, O. Zaïane, H. Hershkovitz, M. Yudelson & J. Stamper (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 168-171).

Rau, M. A., Rummel, N., Aleven, V., Pacilio, L., & Tunc-Pekkan, Z. (2012). 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 Vol 1* (pp. 64-71). International Society of Learning Sciences.

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

Reed, S. K., & Ettinger, M. (1987). Usefulness of tables for solving word problems. *Cognition and Instruction, 4*(1), 43-59.

Reimer, K., & Moyer, P. S. (2005). Third-graders learn about fractions using virtual manipulatives: A classroom study. *Journal of Computers in Mathematics and Science Teaching, 24*(1), 5-25.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science: A Multidisciplinary Journal, 21*(1), 1-29.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction, 12*(5), 529-556.

Renkl, A. (2005). The worked-out example principle in multimedia learning. In R. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 229-246). Cambridge University Press.

Renkl, A., Atkinson, R. K., & Große, C. S. (2004). How Fading Worked Solution Steps Works - A Cognitive Load Perspective. *Instructional Science, 32*(1), 59-82.

Resnick, L. B., & Omanson, S. F. (1987). Learning to understand arithmetic. In R. Glaser (Ed.), *Advances in Instructional Psychology Vol. 3* (pp. 41-95). Hillsdale, NJ: Lawrence Erlbaum Associates.

Riddle, M., & Rodzwell, B. (2000). Fractions: What Happens between Kindergarten and the Army? *Teaching Children Mathematics, 7*(4), 202-206.

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. T. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review, 14*(2), 249-255.

Rittle-Johnson, B., & Koedinger, K. R. (2005). Designing Knowledge Scaffolds to Support Mathematical Problem Solving. *Cognition and Instruction, 23*(3), 313-349.

Rohrer, D. (2008). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education, 40*(1), 4-17.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39*(5), 406–412.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science, 35*, 481-498.

Roussou, M., Oliver, M., & Slater, M. (2006). The virtual playground: an educational virtual reality environment for evaluating interactivity and conceptual learning. *Virtual reality, 10*(3), 227-240.

Salden, R., Aleven, V., Renkl, A., & Schwonke, R. (2008). Worked examples and tutored problem solving: redundant or synergistic forms of support? In C. Schunn (Ed.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 203-213). New York, NY: Lawrence Erlbaum.

Schmidt-Weigand, F., Hänze, M., & Wodzinski, R. (2009). Complex problem solving and worked examples. *Zeitschrift für Pädagogische Psychologie, 23*(2), 129-138.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010a). A closer look at split visual attention in system-and self-paced instruction in multimedia learning. *Learning and Instruction, 20*(2), 100-110.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010b). Explaining the modality and contiguity effects: New insights from investigating students' viewing behaviour. *Applied Cognitive Psychology, 24*(2), 226-237.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207-217.

Schneider, W. (1985). Training High-Performance Skills - Fallacies and Guidelines. *Human Factors, 27*(3), 285-300.

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

Schoor, C., & Bannert, M. (2010). Are pictures more motivating than text? Effects of presentation codality on interest. Paper presented at the EARLI SIG 6-7.

Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (2000). *Cognitive Task Analysis*. Erlbaum Associates.

Schwartz, D. L., & Black, J. B. (1996). Analog imagery in mental model reasoning: Depictive models. *Cognitive Psychology, 30*(2), 154-219.

Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science: A Multidisciplinary Journal*, *20*(4), 457-497.

Schwonke, R., Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology*, *23*, 1227-1243.

Schwonke, R., Ertelt, A., & Renkl, A. (2008). Fostering the translation between external representations. Does it enhance learning with an intelligent tutoring program? In J. Zumbach, N. Schwartz, T. Seufert & L. Kester (Eds.), *Beyond knowledge: The legacy of competence* (pp. 117-119). Springer Netherlands.

Schwonke, R., & Renkl, A. (2010). How do proficient learners construct mental representations of different but related external representations? Paper presented at the EARLI SIG 2.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior*, *25*, 258-266.

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education*, *46*(4), 426-445.

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

Shea, C. H., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(2), 179-187.

Siegler, R. S., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y. et al. (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., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M. et al. (2012). Early Predictors of High School Mathematics Achievement. *Psychological science*, *23*(7), 691-697.

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

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology, 27*(4), 907-912.

Simons, D. J., & Keil, F. C. (1995). An abstract to concrete shift in the development of biological thought: The insides story. *Cognition, 56*(2), 129-163.

Singley, M. K. (1990). The Reification of Goal Structures in a Calculus Tutor: Effects on Problem-Solving Performance. *Interactive Learning Environments, 1*(2), 102-123.

Slykhuis, D. A., Wiebe, E. N., & Annetta, L. A. (2005). Eye-tracking students' attention to PowerPoint photographs in a science education setting. *Journal of Science Education and Technology, 14*(5), 509-520.

Smith, L. B. (2003). Learning to recognize objects. *Psychological Science, 14*(3), 244-250.

Solaz-Portolés, J. J., & Lopez, V. S. (2007). Presentations in problem solving in science: Directions for practice. *Asia-Pacific Forum on Science Learning and Teaching, 8*(2).

Soloway, E., Jackson, S. L., Klein, J., Quintana, C., Reed, J., Spitulnik, J. et al. (1996). Learning theory in practice: case studies of learner-centered design. In *Proceedings of the SIG CHI Conference on Human Factors in Computing Systems* (pp. 189-196). ACM Press.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology, 31*(1), 24-33.

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

Stafylidou, S., & Vosniadou, S. (2004). The development of students' understanding of the numerical value of fractions. *Learning and Instruction, 14*(5), 503-518.

Steencken, E. P., & Maher, C. (2002). *Young Children's Growing Understanding of Fraction Ideas Making Sense of Fractions, Ratios, and Proportions*. Reston, VA: National Council of Teachers of Mathematics.

Steiner, G. F., & Stoecklin, M. (1997). Fraction calculation--a didactic approach to constructing mathematical networks. *Learning and Instruction, 7*(3), 211-233.

Stieff, M., Hegarty, M., & Deslongchamps, G. (2011). Identifying representational competence with multi-representational displays. *Cognition and Instruction, 29*(1), 123-145.

Suh, J., & Moyer, P. S. (2007). Developing Students' Representational Fluency Using Virtual and Physical Algebra Balances. *Journal of Computers in Mathematics and Science Teaching, 26*(2), 155-173.

Suh, J., Moyer, P. S., & Heo, J.-J. (2005). Examining technology uses in the classroom: Developing fraction sense using virtual manipulative concept tutorials. *Journal of Interactive Online Learning, 3*(4), 1-21.

Superfine, A. C., Canty, R. S., & Marshall, A. M. (2009). Translation between external representation systems in mathematics: All-or-none or skill conglomerate? *The Journal of Mathematical Behavior*, *28*(4), 217-236.

Suthers, D., Vatrapu, R., Medina, R., Joseph, S., & Dwyer, N. (2008). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers & Education, 50*(4), 1103-1127.

Sweller, J. (2006). The worked example effect and human cognition. Lea*rning and Instruction, 16*, 165-169.

Tabachneck, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. *Cognitive Science*, *21*(3), 305-350.

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

Tatsuoka, K. K. (1984). *Analysis of Errors in Fraction Addition and Subtraction Problems.* Access ERIC: FullText.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*, 837-848.

Test, D. W., & Ellis, M. F. (2005). The Effects of LAP Fractions on Addition and Subtraction of Fractions with Students with Mild Disabilities. *Education & Treatment of Children, 28*(1), 11-24.

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review, 14*(4), 663-668.

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 of Mathematics.

Thußbas, C. (2001). Beeinflussen Cover Stories mathematischer Texaufgaben analogen Transfer in der Mathematik? *Zeitschrift für Experimentelle Psychologie, 48*(4), 272-289.

Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2005). Simple Addition Strategies in a First-Grade Class With Multiple Strategy Instruction. *Cognition and Instruction*, *23*(1), 1-21.

Tsui, C.-Y., & Treagust, D. F. (2013). Introduction to multiple representations: Their importance in biology and biological education. In D. F. Treagust & C.-Y. Tsui (Eds.), *Multiple Representations in Biological Education* (pp. 3-18). Springer Netherlands.

Tunc-Pekkan, Z., Rau, M., Aleven, V., & Rummel, N. (2010). External Representations and Fractional Knowledge. Paper presented at the Third Annual inter-Science of Learning Center (iSLC) Conference For Students and Postdoctoral Fellows at the Science of Learning Centers.

Tunç-Pekkan, Z., Zeylikman, L., V., A., & Rummel, N. (2010). Fifth Graders' Conception of Fractions on Numberline Representations. Paper presented at the annual meeting of North American Chapter of the International Group for the Psychology of Mathematics Education.

Urban-Woldron, J. (2009). Interactive simulations for the effective learning of physics. *Journal of Computers in Mathematics and Science Teaching, 28*(2), 163-176.

Uttal, D. H. (2003). On the relation between play and symbolic thought: The case of mathematics manipulatives. In O. Saracho & B. Spodek (Eds.), *Contemporary Perspectives in Early Childhood* (pp. 97-114). Information Age Press.

van der Meij, J. (2007). *Support for learning with multiple representations designing simulation-based learning environments*. Universiteit Twente, Netherlands.

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.

van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*(1), 16-26.

van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction*, *20*(2), 95-99.

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

van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. G. W. C. (2002). Redirecting learners' attention during training: effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction*, *12*(1), 11-37.

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.

Vanhille, L. S., & Baroody, A. J. (2002). Fraction Instruction That Fosters Multiplicative Reasoning. In B. Litwiller & G. Bright (Eds.), *Making sense of fractions, ratios, and proportions* (pp. 224-236). Reston, VA: National Council of Teachers of Mathematics.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educational Psychologist, 46*(4), 197-221.

Walkington, C., Nathan, M., Wolfgram, M., Alibali, M., & Srisurichan, R. (2011). Bridges and barriers to constructing conceptual cohesion across modalities and temporalities: Challenges of STEM integration in the precollege engineering classroom. In S. Publishers (Ed.), *Engineering in PreCollege Settings: Research into Practice*.

Wells, J., & Lewis, L. (2006). *Internet access in US public schools and classrooms: 1994-2005*. National Center for Education Statistics.

Witherspoon, M. L. (1993). Fractions: In Search of Meaning. *Arithmetic Teacher, 40*(8), 482-485.

Wong, R. M. F., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction, 12*(2), 233-262.

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

Yanik, H. B., Helding, B., & Flores, A. (2008). Teaching the concept of unit in measurement interpretation of rational numbers. *Elementary Education Online*, *7*(3), 693-705

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, *21*(2), 179-217.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. Cognitive *Science: A Multidisciplinary Journal*, *18*(1), 87-122.

Zhang, Z. H., & Linn, M. C. (2011). Can generating representations enhance learning with dynamic visualizations? *Journal of Research in Science Teaching*, *48*(10), 1177-1198.

Appendix 1: Fractions Tutor curriculum

Topic	Problem type / Learning goal	Standards
Topic 1: Naming Fractions		
	Problem type: Determine what fraction (unit fractions and proper fractions) a given graphical representation shows (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Become familiar with the (interactive) graphical representations and link it to the symbolic representation	PA State Standard 2.1.5.D[9] NCTM Standards Grades 3-5 #1[10] NCTM Standards Grades 3-5 #3 Common Core Standard 3.NF.1[11] Common Core Standard 3.NF.2 Common Core Standard 3.NF.3
	Problem type: Compare fractions given a graphical representation and the corresponding symbolic fraction (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Developing a sense for the size of fractions using the graphical representation and link it to the symbolic representation	PA State Standard 2.1.5.D PA State Standard 2.4.5.A PA State Standard 2.4.5.B NCTM Standards Grades 3-5 #1 NCTM Standards Grades 3-5 #3 Common Core Standard 3.NF.1 Common Core Standard 3.NF.2 Common Core Standard 3.NF.3
	Problem type: Determine what fraction a given graphical representation represents, given the unit of the fraction (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Understand that the unit determines the relative size of the fraction	PA State Standard 2.1.3.B PA State Standard 2.3.3.B PA State Standard 2.1.5.D NCTM Standards Grades 3-5 #1 NCTM Standards Grades 3-5 #3 Common Core Standard 3.NF.1 Common Core Standard 3.NF.2 Common Core Standard 3.NF.3
Topic 2: Making Graphical Representations of Fractions		
	Problem type: Construct a graphical representation for a fraction given symbolically (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Become familiar with the (interactive) graphical representations and link it to the symbolic representation	PA State Standard 2.1.5.D NCTM Standards Grades 3-5 #1 NCTM Standards Grades 3-5 #3 Common Core Standard 3.NF.1 Common Core Standard 3.NF.2 Common Core Standard 3.NF.3
	Problem type: Compare fractions given a graphical representation and the corresponding symbolic fraction (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Developing a sense for the size of fractions using the graphical representation and link it to the symbolic representation	PA State Standard 2.1.5.D PA State Standard 2.4.5.A PA State Standard 2.4.5.B NCTM Standards Grades 3-5 #1 NCTM Standards Grades 3-5 #3 Common Core Standard 3.NF.1 Common Core Standard 3.NF.2 Common Core Standard 3.NF.3
	Problem type: Construct a graphical representation for a fraction given symbolically while using different units (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Understand that the unit determines the relative size of the fraction	PA State Standard 2.1.3.B PA State Standard 2.3.3.B PA State Standard 2.1.5.D NCTM Standards Grades 3-5 #1 NCTM Standards Grades 3-5 #3 Common Core Standard 3.NF.1

[9] see Appendix 2 for Pennsylvania State Standard definitions
[10] see Appendix 3 for NCTM Standard definitions
[11] see Appendix 4 for Common Core Standard definitions

		Common Core Standard 3.NF.2 Common Core Standard 3.NF.3
Topic 3: Reconstructing The Unit		
	Problem type: Given a unit fraction, reconstruct the unit of the fraction (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Understand that the unit determines the relative size of the fraction, developing a sense for the size of fractions using the graphical representation and link it to the symbolic representation	PA State Standard 2.1.3.B PA State Standard 2.3.3.B PA State Standard 2.1.5.D Common Core Standard 3.NF.1 Common Core Standard 3.NF.2
	Problem type: Given a proper fraction, (a) find the unit fraction, and (b) reconstruct the unit of the fraction (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Understand that the unit determines the relative size of the fraction, developing a sense for the size of fractions using the graphical representation and link it to the symbolic representation	PA State Standard 2.1.3.B PA State Standard 2.3.3.B PA State Standard 2.1.5.D Common Core Standard 3.NF.1 Common Core Standard 3.NF.2
Topic 4: Naming Improper Fractions		
	Problem type: Determine what improper fraction a given graphical representation shows (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Understand that fractions can be larger than 1	PA State Standard 2.1.5.D NCTM Standards Grades 3-5 #1 Common Core Standard 3.NF.1 Common Core Standard 3.NF.2
	Problem type: Determine what fraction a given graphical representation represents, given the unit of the fraction (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Understand that the unit determines the relative size of the fraction	PA State Standard 2.1.3.B PA State Standard 2.3.3.B PA State Standard 2.1.5.D NCTM Standards Grades 3-5 #1 Common Core Standard 3.NF.1 Common Core Standard 3.NF.2
Topic 5: Making Graphical Representations of Improper Fractions		
	Problem type: Construct a graphical representation for a fraction given symbolically (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Become familiar with the (interactive) graphical representations and link it to the symbolic representation	PA State Standard 2.1.5.D NCTM Standards Grades 3-5 #1 Common Core Standard 3.NF.1 Common Core Standard 3.NF.2
	Problem type: Compare fractions given a graphical representation and the corresponding symbolic fraction (using circle diagrams, rectangles, and number lines) ***Learning goal:*** Developing a sense for the size of fractions using the graphical representation and link it to the symbolic representation	PA State Standard 2.1.5.D PA State Standard 2.4.5.A PA State Standard 2.4.5.B NCTM Standards Grades 3-5 #1 Common Core Standard 3.NF.1 Common Core Standard 3.NF.2
Topic 6: Equivalent Fractions: Underlying Concepts		
	Problem type: Given a graphical representation of a fraction (circle diagrams, rectangles, and number lines), manipulate it to find an equivalent fraction graphically, and name the corresponding symbolic fraction ***Learning goal:*** Understand the invariance of amounts when partitioning a fraction into more sections	PA State Standard 2.1.8.A NCTM Standards Grades 3-5 #1 Common Core Standard 4.NF.1*** Common Core Standard 3.NF.3
	Problem type: Given a symbolic fraction, manipulate numerator and denominator separately, while observing corresponding changes in a graphical representation (circle diagrams, rectangles, and number lines) ***Learning goal:*** Understand that multiplying numerator and denominator by	PA State Standard 2.1.8.A NCTM Standards Grades 3-5 #1 Common Core Standard 4.NF.1 Common Core Standard 3.NF.3

	the same number does not change the amount of the fraction	
Topic 7: Equivalent Fractions: Expanding and Reducing		
	Problem type: Given a graphical representation of a unit fraction (circle diagrams, rectangles, and number lines), manipulate it to expand it and reduce it ***Learning goal:*** Understand that expanding and reducing fractions are interchangeable activities	PA State Standard 2.1.8.A NCTM Standards Grades 3-5 #1 Common Core Standard 3.NF.3 Common Core Standard 4.NF.1
	Problem type: Given a graphical representation of a proper fraction (circle diagrams, rectangles, and number lines), manipulate it to expand it and reduce it ***Learning goal:*** Understand that expanding and reducing fractions are interchangeable activities	PA State Standard 2.1.8.A NCTM Standards Grades 3-5 #1 Common Core Standard 3.NF.3 Common Core Standard 4.NF.1
Topic 8: Comparing Fractions		
	Problem type: Given two fractions and their graphical representation (circle diagrams, rectangles, or number lines), use common benchmarks (e.g., ½, ¼) to compare them ***Learning goal:*** Being able to use benchmarks to compare fractions	PA State Standard 2.2.8.B NCTM Standards Grades 3-5 #1 NCTM Standards Grades 3-5 #2 Common Core Standard 3.NF.3 Common Core Standard 4.NF.2
	Problem type: Given two fractions and their graphical representation (circle diagrams, rectangles, or number lines), use equivalent fractions to convert them to the same numerator or denominator to reason that one is larger than the other ***Learning goal:*** Being able to use equivalent fractions to compare fractions	PA State Standard 2.2.8.B NCTM Standards Grades 3-5 #1 NCTM Standards Grades 3-5 #2 Common Core Standard 3.NF.3 Common Core Standard 4.NF.2
Topic 9: Adding Fractions		
	Problem type: Add two given fractions with the same denominators and specify the unit of the addend fractions and the sum fraction ***Learning goal:*** Understanding that the unit does not change when adding two fractions, and that for this reason, the denominator of the sum fraction remains the same	PA State Standard 2.2.8.B Common Core Standard 4.NF.3
	Problem type: Add two given fractions with the different denominators and specify the unit of the addend fractions and the sum fraction ***Learning goal:*** Understanding that the unit does not change when adding two fractions as the motivation for finding the common denominator before adding fractions; being able to use equivalent fractions to find the common denominator of two addend fractions	PA State Standard 2.2.8.B Common Core Standard 5.NF.1
Topic 10: Subtracting Fractions		
	Problem type: Subtract two given fractions with the same denominators and specify the unit of the subtrahend fractions and the difference fraction ***Learning goal:*** Understanding that the unit does not change when subtracting two fractions, and that for this reason, the denominator of the difference fraction remains the same	PA State Standard 2.2.8.B Common Core Standard 4.NF.3
	Problem type: Subtract two given fractions with the different denominators and specify the unit of the subtrahend fractions and the difference fraction ***Learning goal:*** Understanding that the unit does not change when subtracting two fractions as the motivation for finding the common denominator before subtracting fractions; being able to use equivalent fractions to find	PA State Standard 2.2.8.B Common Core Standard 5.NF.1

	the common denominator of two subtrahend fractions	

Table A1. Topics and problem types covered by the Fractions Tutor.

Appendix 2: Pennsylvania State Standards

Pennsylvania State Standards	**Alignment of Fractions Tutor with the Standards**
Use whole numbers and fractions to represent quantities. (2.1.3. B) Use drawings, diagrams or models to show the concept of fraction as part of a whole. (2.1.3. D)	In the earlier units of the Fractions Tutor, students will work with multiple graphical representations of fractions and will have to translate between the graphical representations and the symbolic representations of fractions. We will help students transition from counting shaded sections and total sections in circles and rectangles to viewing fractions as ordered quantities in number lines.
Use models to represent fractions and decimals (2.1.5. D)	Circles, rectangles, and number lines are used throughout the Fractions Tutor units
Develop and apply algorithms to solve word problems that involve addition, subtraction, and/or multiplication with fractions and mixed numbers that include like and unlike denominators. (2.2.5. C)	Specific units in the Fractions Tutor focus on addition, subtraction, multiplication and division with fractions and mixed numbers, using graphical representations to support students' conceptual understanding of algorithms. The Fractions Tutor uses realistic cover stories to introduce the graphical representations.

Table A2. Alignment of the Fractions Tutor with Pennsylvania State Standards.

Appendix 3: NCTM Standards

NCTM Standards	Alignment of Fractions Tutor with the Standards
All students should develop understanding of fractions as parts of unit wholes, as parts of a collection, as locations on number lines, and as divisions of whole numbers.	The Fractions Tutor uses graphical representations to illustrate the part-whole interpretation of fractions (area models: circle and rectangle), and the measurement interpretation (number line).
All students should use models, benchmarks, and equivalent forms to judge the size of fractions.	The Fractions Tutor uses graphical models to illustrate the relative size of fractions and to support fraction comparison. Graphical models are also used to define equivalent fractions as fractions that show the same amount using different numerators and denominators.
All students should explore numbers less than 0 by extending the number line and through familiar applications.	The Fractions Tutor provides practice in interpreting and manipulating fractions using circles, rectangles, and number lines.

Table A3. Alignment of the Fractions Tutor with Pennsylvania State Standards.

Appendix 4: Common Core Standards

Grade	Common Core Standards	Alignment of Fractions Tutor with the Standards
3	Understand a fraction 1/*b* as the quantity formed by 1 part when *a* whole is partitioned into *b* equal parts; understand a fraction *a*/*b* as the quantity formed by a parts of size 1/*b*. (3.NF.1.)	Throughout the early units of the fractions tutor, we emphasize the unit of the fraction as what the fraction is being taken of. The fractions tutor includes partitioning activities and repetition activities of unit fractions to form proper fractions using circles, rectangles, and number lines.
	Understand a fraction as a number on the number line; represent fractions on a number line diagram. (3.NF.2.) Represent a fraction 1/*b* on a number line diagram by defining the interval from 0 to 1 as the whole and partitioning it into *b* equal parts. Recognize that each part has size 1/*b* and that the endpoint of the part based at 0 locates the number 1/*b* on the number line. Represent a fraction *a*/*b* on a number line diagram by marking off a lengths 1/*b* from 0. Recognize that the resulting interval has size *a*/*b* and that its endpoint locates the number *a*/*b* on the number line.	The fractions tutor includes activities with the number line while supporting students in making connections between circles, rectangles, and number lines. The fractions tutor includes number lines that extend beyond 1 even when showing fractions between 0 and 1. Throughout the number line activities the fractions tutor emphasizes that the unit of a fraction on the number line is the distance between 0 and 1. The fractions tutor also includes reflection questions to help students understand that a proper fraction can be constructed by repeating unit fractions.
	Explain equivalence of fractions in special cases, and compare fractions by reasoning about their size. (3.NF.3.) Understand two fractions as equivalent (equal) if they are the same size, or the same point on a number line. Recognize and generate simple equivalent fractions, e.g., 1/2 = 2/4, 4/6 = 2/3). Explain why the fractions are equivalent, e.g., by using a visual fraction model. Express whole numbers as fractions, and recognize fractions that are equivalent to whole numbers. Compare two fractions with the same numerator or the same denominator by reasoning about their size. Recognize that comparisons are valid only when the two fractions refer to the same whole. Record the results of comparisons with the symbols >, =, or <, and justify the conclusions, e.g., by using a visual fraction model.	Throughout the early units of the fractions tutor, we explicitly ask students to reason about the size of two fractions that have either the same denominator but different numerators, or different denominators but the same numerators. We use circles, rectangles, and number lines to support their thinking. The fractions tutor includes two units on equivalent fractions where we first introduce equivalent fractions conceptually and then provide computational practice in finding equivalent fractions.

4	Explain why a fraction *a*/*b* is equivalent to a fraction ($n \times a$)/($n \times b$) by using visual fraction models, with attention to how the number and size of the parts differ even though the two fractions themselves are the same size. Use this principle to recognize and generate equivalent fractions. (4.NF.1.)	The Fractions Tutor uses graphical re-partitioning of graphical representations without changing the shown amount in order to introduce equivalent fractions, and to demonstrate why numerator and denominator need to be multiplied by the same number in order to conserve the amount.
	Compare two fractions with different numerators and different denominators, e.g., by creating common denominators or numerators, or by comparing to a benchmark fraction such as 1/2. Recognize that comparisons are valid only when the two fractions refer to the same whole. Record the results of comparisons with symbols >, =, or <, and justify the conclusions, e.g., by using a visual fraction model. (4.NF.2.)	The Fractions Tutor introduces fractions with activities in which students are asked to name fractions given a graphical representation. As part of these activities, students will name two fractions and then compare them to one another. In these activities, the unit of the fraction is also being varied. Later in the tutor curriculum, an entire unit is dedicated to fractions comparison, using 1/2 as a benchmark.
	Understand a fraction *a*/*b* with $a > 1$ as a sum of fractions 1/*b*. (4.NF.3.) Understand addition and subtraction of fractions as joining and separating parts referring to the same whole. Decompose a fraction into a sum of fractions with the same denominator in more than one way, recording each decomposition by an equation. Justify decompositions, e.g., by using a visual fraction model. Add and subtract mixed numbers with like denominators, e.g., by replacing each mixed number with an equivalent fraction, and/or by using properties of operations and the relationship between addition and subtraction. Solve word problems involving addition and subtraction of fractions referring to the same whole and having like denominators, e.g., by using visual fraction models and equations to represent the problem.	The Fractions Tutor includes units on fraction addition and fraction subtraction in which proper fractions will be decomposed into unit fractions. The concept of the unit of the fraction will be emphasized throughout the entire curriculum, and (wrt fraction addition and subtraction) special emphasis will be put on helping students understand that the denominator defines the size of the sections that are being added (in relation to the unit) and should therefore remain the same (i.e., adding the numerators, but not the denominators). The addition and subtraction units of the Fractions Tutor will also include mixed numbers.
5	Add and subtract fractions with unlike denominators (including mixed numbers) by replacing given fractions with equivalent fractions in such a way as to produce an equivalent sum or difference of fractions with like denominators. (5.NF.1.)	The Fractions Tutor includes fraction addition and fraction subtraction problems in which students are first guided to convert fractions to that they have the least common denominator.

Table A4. Alignment of the Fractions Tutor with Common Core Standards.

Appendix 5: Experiment 2 test items

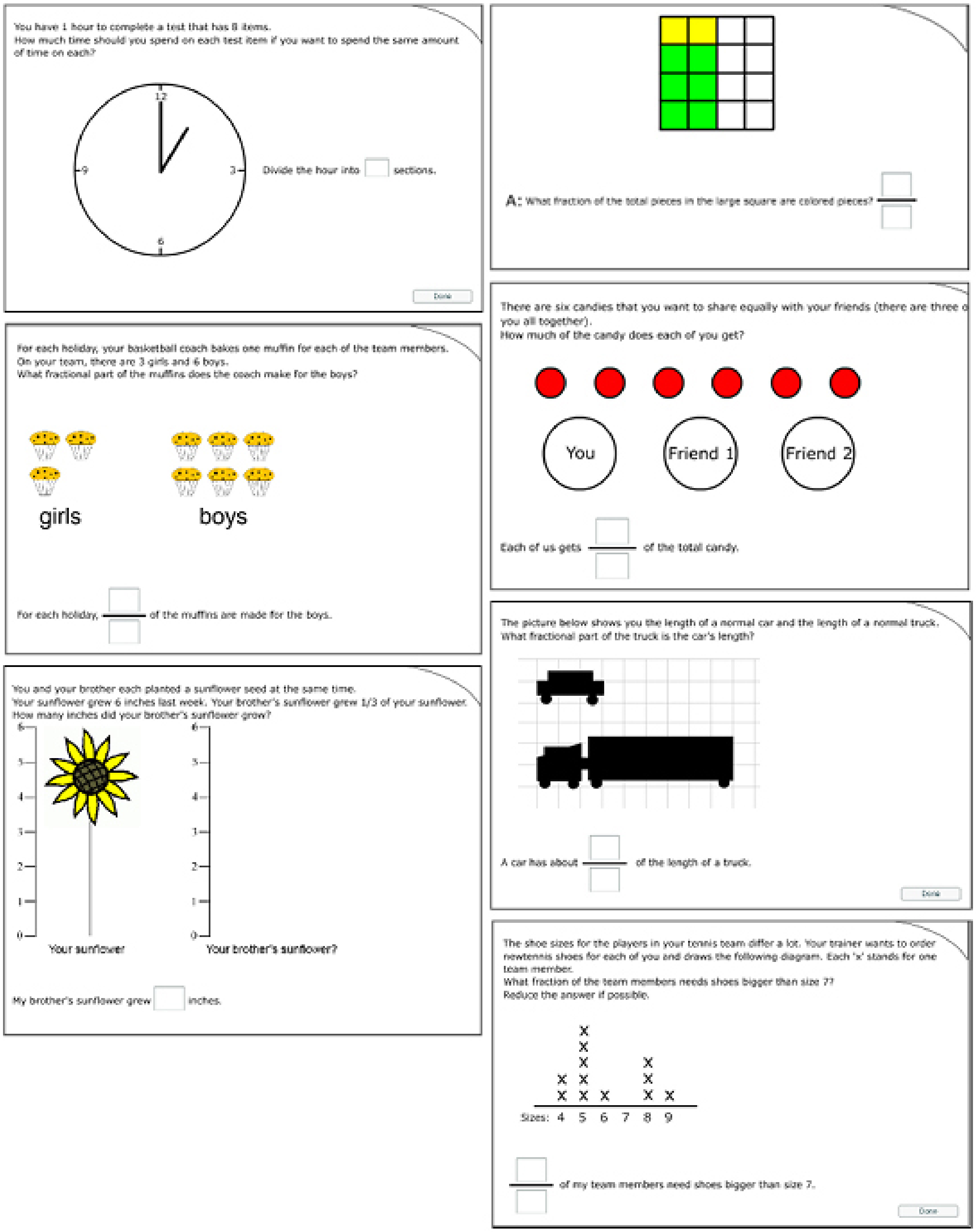


Fig. A5-1. Example test items from the representational knowledge test used in Experiment 2.

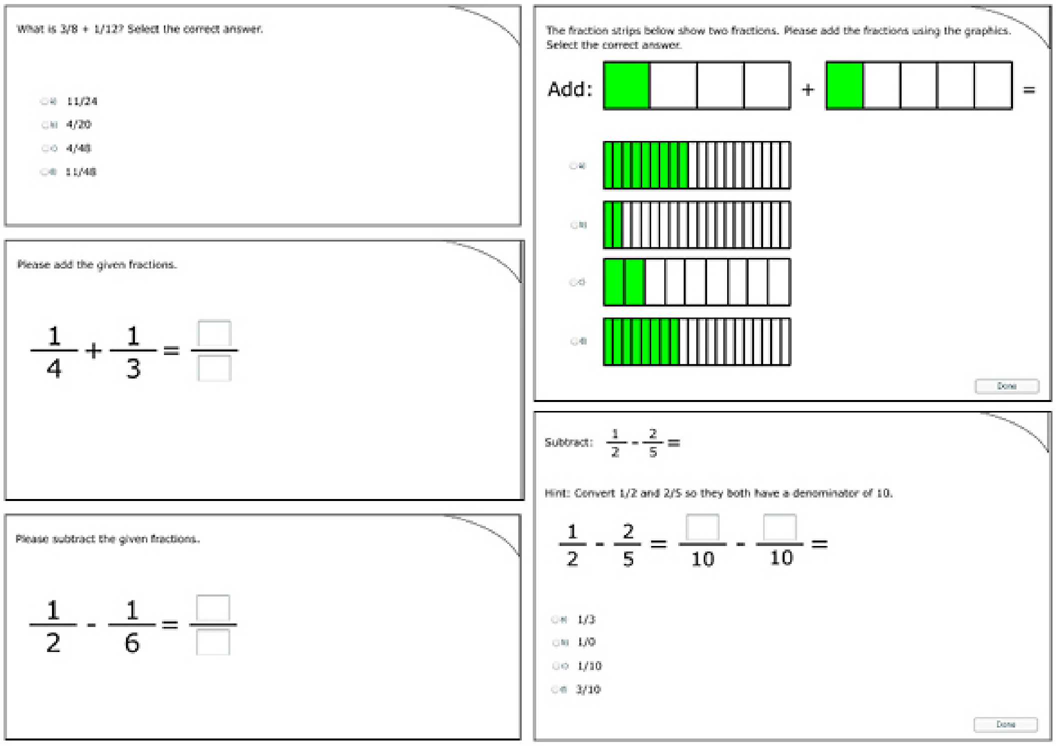


Fig. A5-2. Example test items from the operational knowledge test used in Experiment 2.

Appendix 6: Experiment 3 test items

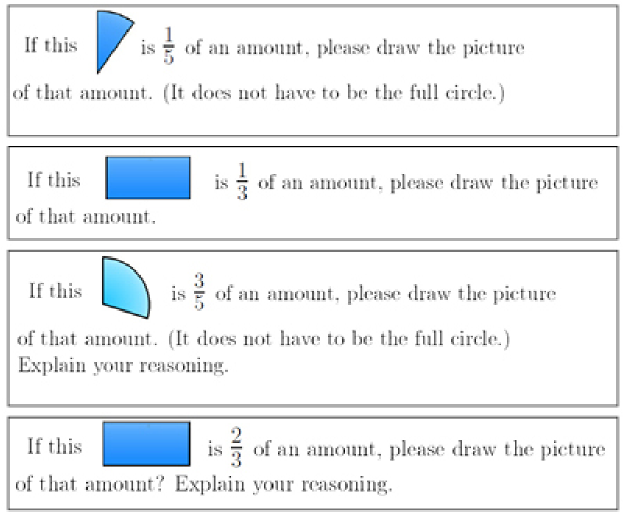


Fig. A6-1. Example test items from the area models fluency test used in Experiment 3.

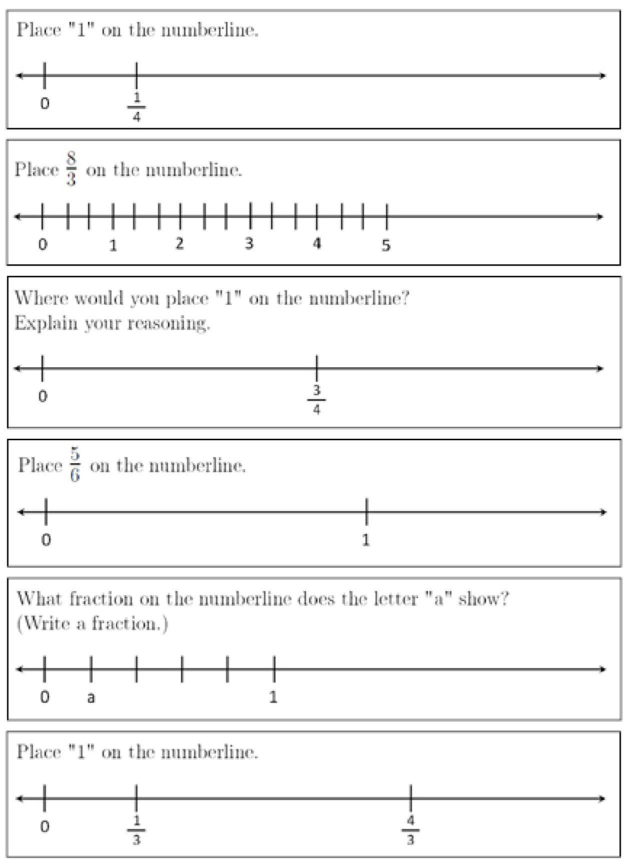


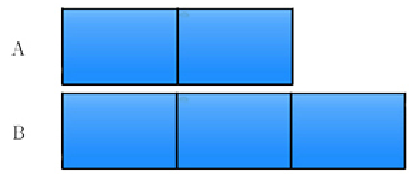
Fig. A6-2. Example test items from the number lines fluency test used in Experiment 3.

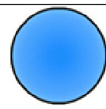
There is a bag of apples. If 9 apples are $\frac{1}{3}$ of all the apples in the bag, how many apples are in the bag?

Joe's weekly allowance is 12 dollars. How much is $\frac{1}{3}$ of his allowance?

There is a bag of candy pieces. If 12 candies are $\frac{2}{3}$ of all the candies in the bag, how many candies are in the bag?

If Rectangle A (upper rectangle) is the unit, what fraction does Rectangle B (bottom rectangle) show?



If this  is the unit, what would be the fraction shown by the shaded parts of the diagrams below? (Write as an improper fraction or as a mixed number.)

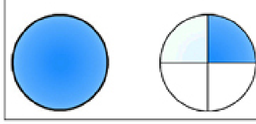
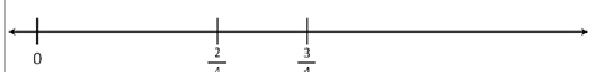


Fig. A6-3. Example test items from the conceptual transfer test used in Experiment 3.

Using the numberline, find a fraction that is larger than $\frac{2}{4}$ and smaller than $\frac{3}{4}$ and place it on the numberline.
(You may change the number of sections on the numberline.)



Compare $\frac{7}{6}$ and $\frac{11}{7}$. Which fraction is larger?
Explain your thinking with pictures or words.

Compare $\frac{4}{7}$ and $\frac{5}{8}$. Which fraction is larger?
Explain your thinking with pictures or words.

Fig. A6-4. Example test items from the procedural transfer test used in Experiment 3.

Appendix 7: Experiment 4 test items

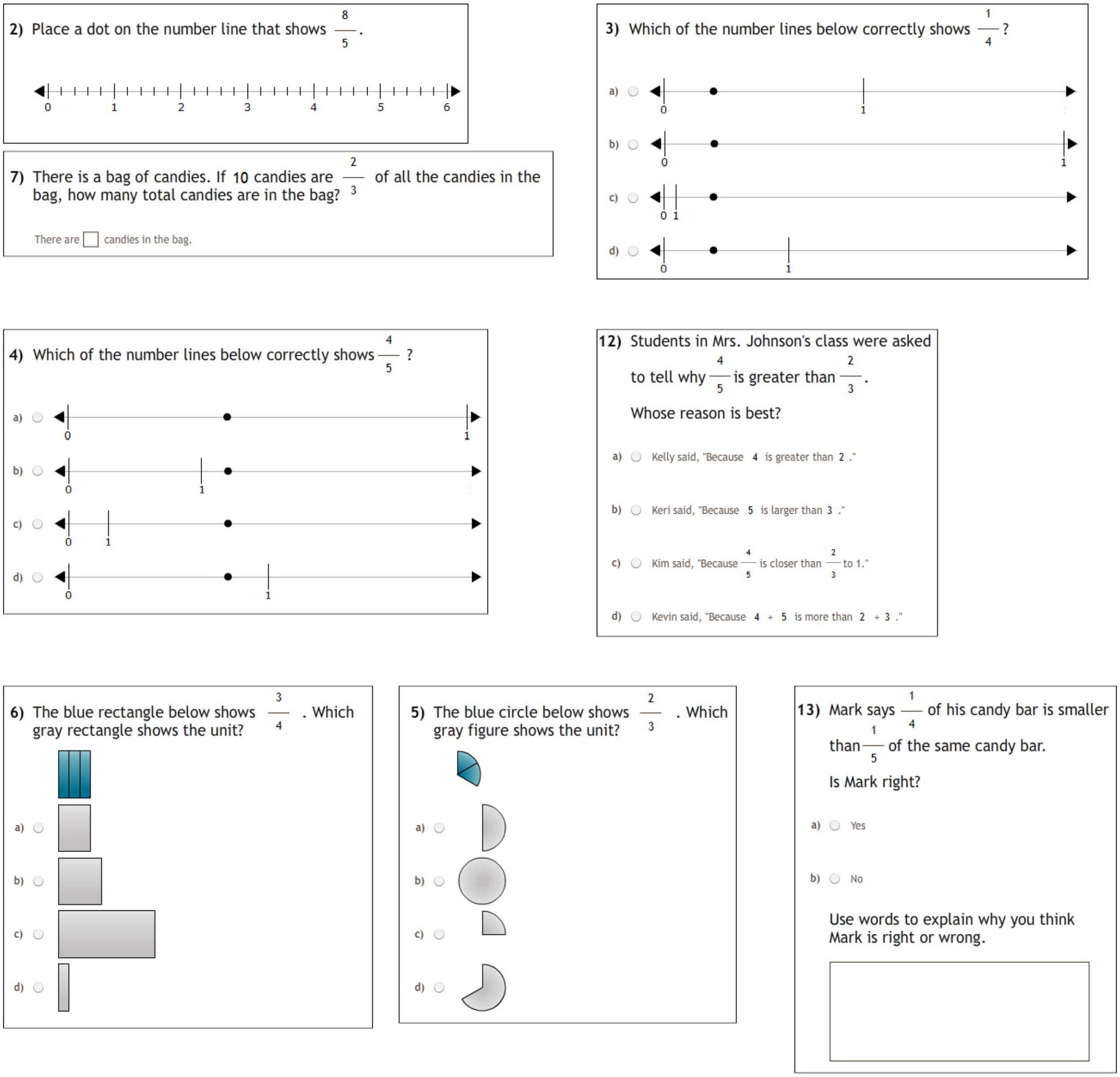


Fig. A7-1. Example test items from the conceptual knowledge test used in Experiment 4.

Fig. A7-2. Example test items from the procedural knowledge test used in Experiment 4.

Appendix 8: Experiment 5 test items

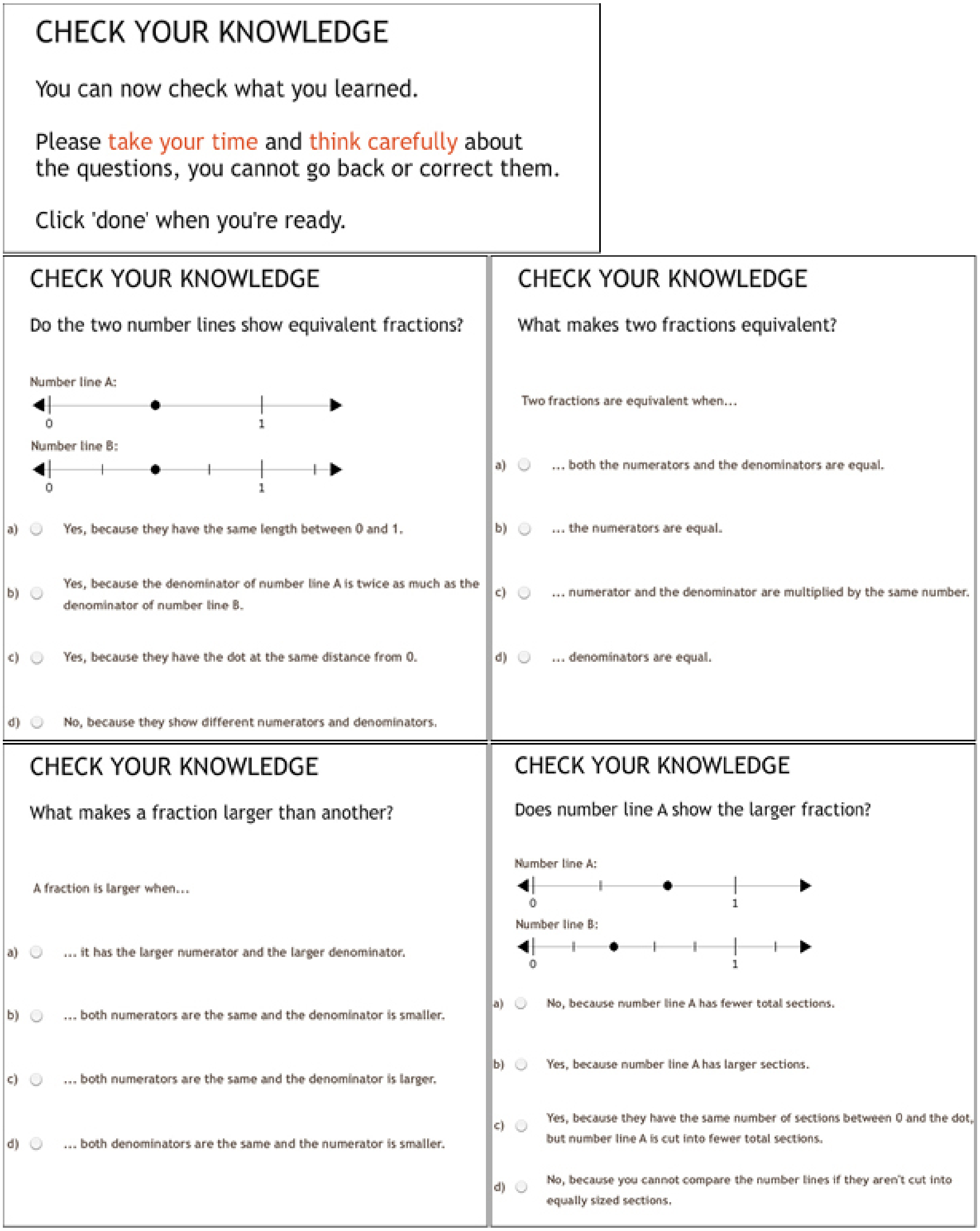


Fig. A8-1. Example test items from the sense-making quiz used in Experiment 5.

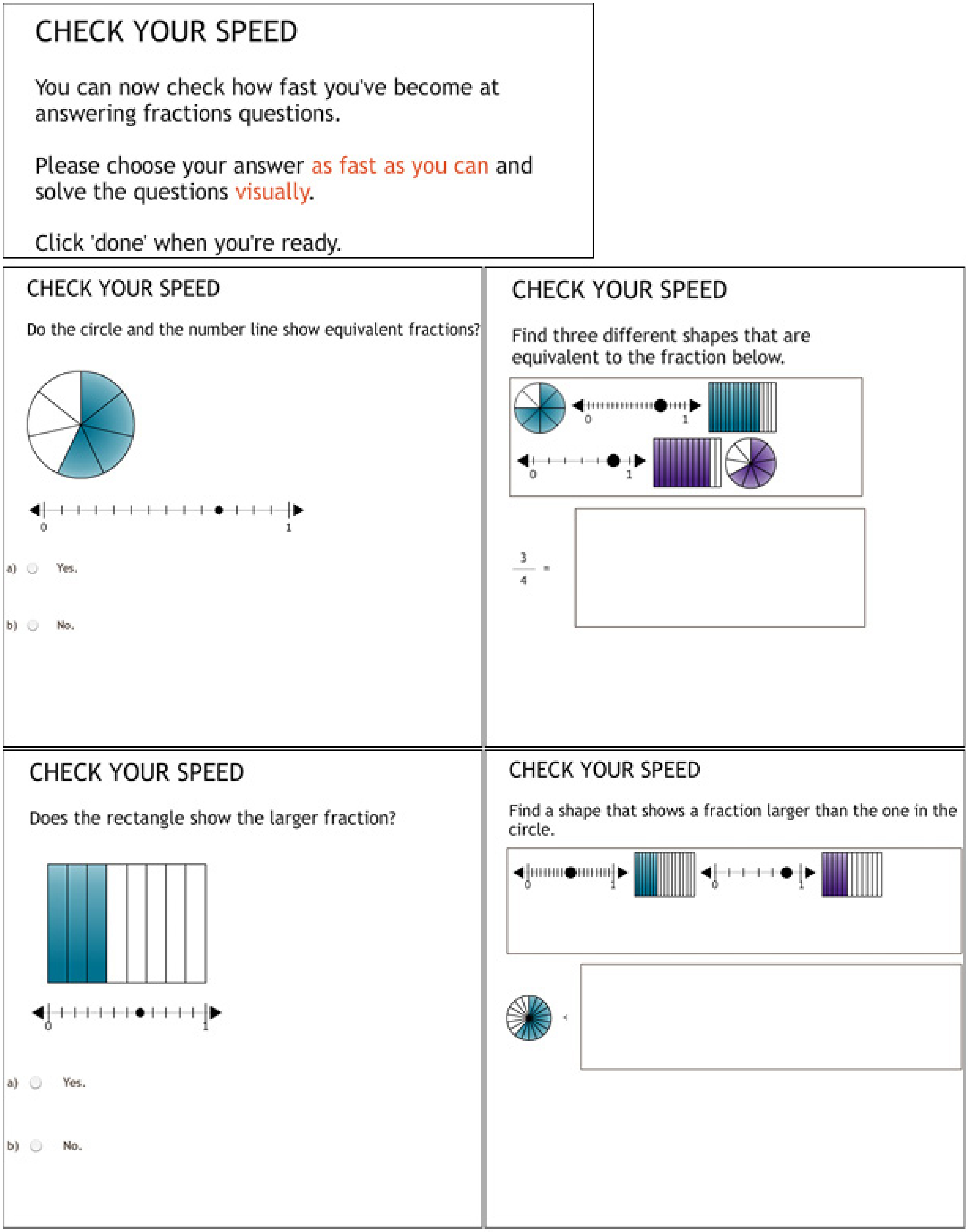


Fig. A8-2. Example test items from the fluency-building quiz used in Experiment 5.

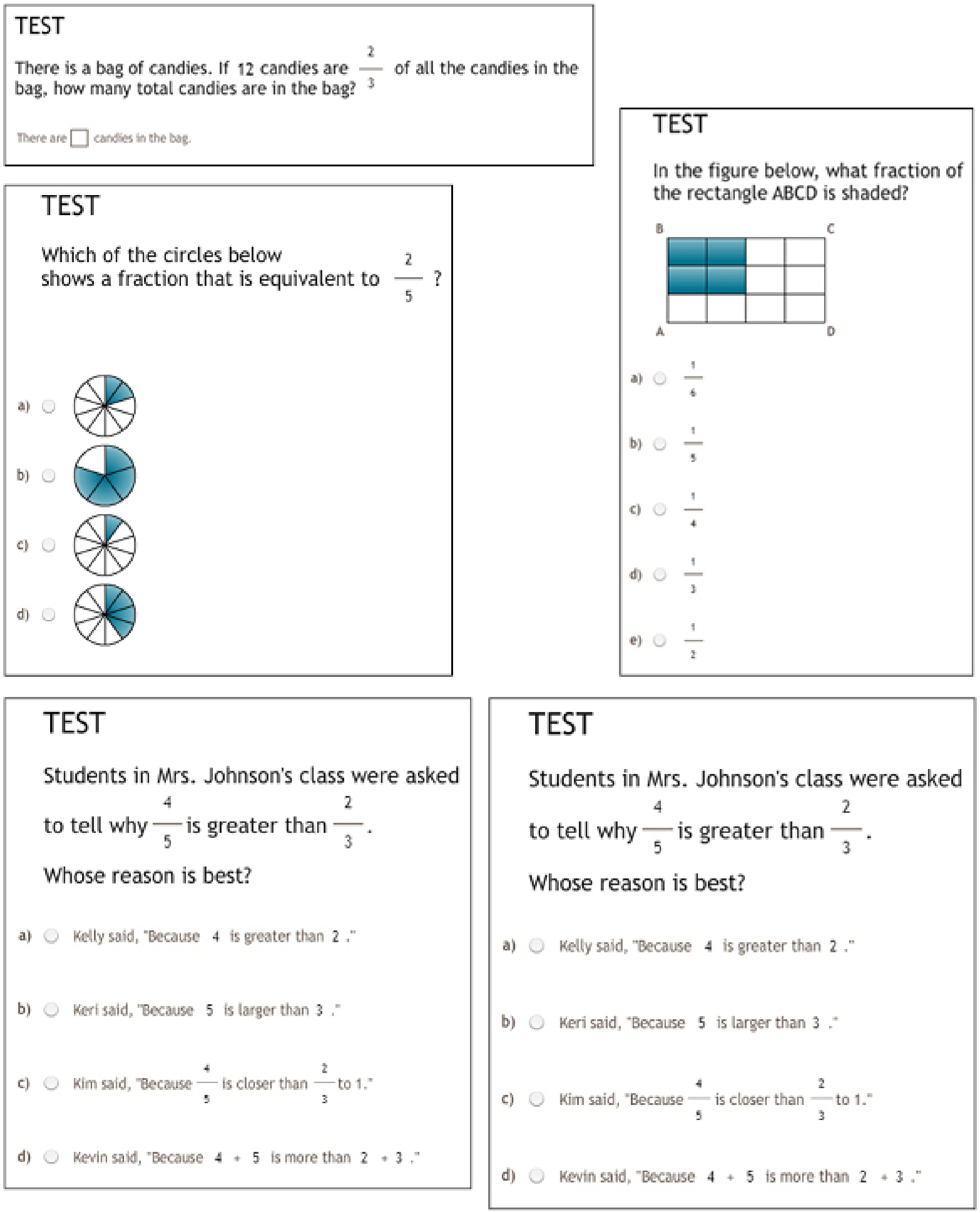


Fig. A8-3. Example test items from the transfer test used in Experiment 5.