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

Martina Rau

Ph.D. Thesis
Human-Computer Interaction Institute
Carnegie Mellon University

Ph.D. Committee:
Vincent Aleven, Human-Computer Interaction Institute, Carnegie Mellon University
Nikol Rummel, Institute of Educational Research, Ruhr-Universität Bochum
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
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. But to benefit from multiple graphical representations, students need to accomplish a number of cognitive tasks. They 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. 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.

I provide a theoretical framework that describes the learning processes students have to engage in to benefit from 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 guided the design of an intelligent tutoring system for fractions which was iteratively improved on the basis of empirical evaluations and user-centered techniques. At the same time, the framework served as the theoretical basis for five classroom experiments and lab studies with over 3,000 4th- and 5th-grade students. Each experiment tested predictions made by the theoretical framework, evaluated the effects of different types of instructional support for the proposed learning processes, and served to iteratively improve the Fractions Tutor.

One outcome of my research is an effective educational technology that enhances students’ robust learning of fractions and that is usable within the context of real educational settings. Furthermore, my research provides an empirically validated theoretical framework for learning with multiple graphical representations which addresses shortcomings of prior research on learning with multiple representations. Finally, I provide a set of empirically tested instructional design principles for the support of learning processes that students need to engage in to benefit from multiple graphical representations. By integrating the learning sciences perspective with intelli-
gent tutoring systems and human-computer interaction research, my work yields theoretically motivated, usable, and empirically tested principles for the optimal use of multiple graphical representations in educational technologies.
Table of Contents

1 Introduction ...................................................................................................................................... 8
  1.1 Learning sciences perspective ............................................................................................... 13
  1.2 The educational technology perspective ............................................................................... 17
  1.3 Summary ............................................................................................................................... 20
2 Theoretical Framework .................................................................................................................. 21
  2.1 Existing theoretical frameworks for learning with dual representations ............................... 22
  2.2 A new theoretical framework for learning with multiple graphical representations .......... 26
3 Fractions Tutor ............................................................................................................................... 30
  3.1 Multiple graphical representations to help students learn fractions ...................................... 31
  3.2 Use of multiple, interactive, abstract graphical representations ............................................ 34
  3.3 User-centered design ............................................................................................................. 37
    3.3.1 Cognitive Tutor design process .................................................................................... 37
    3.3.1.1. Stage 1: Stakeholder and problem identification .................................................... 37
    3.3.1.2. Stage 2: Identifying assessment and practice problems ........................................... 38
    3.3.1.3. Stage 3: Cognitive task analysis ............................................................................. 38
    3.3.1.4. Stage 4: Cognitive modeling and tutor development ............................................... 39
    3.3.1.5. Stage 5: Pilot studies and parametric studies ......................................................... 39
    3.3.1.6. Stage 6: Classroom use and evaluation ............................................................... 40
    3.3.2 Identifying stakeholder goals and instructional design principles ................................ 41
    3.3.2.1. Forming a goal hierarchy ....................................................................................... 41
    3.3.3 Resolving conflicts ....................................................................................................... 44
      3.3.3.1. Goal hierarchy ....................................................................................................... 44
      3.3.3.2. Parametric experiments ....................................................................................... 46
      3.3.3.3. Cross-iteration studies ....................................................................................... 46
  3.4 Overview of curriculum .......................................................................................................... 50
    3.4.1 Individual-representation problems for representational understanding and fluency .. 54
    3.4.2 Connectional sense-making support for connectional understanding .......................... 54
    3.4.3 Connectional fluency-building support for connectional fluency ................................ 57
  3.5 Effectiveness ......................................................................................................................... 59
4 Classroom Experiments and Lab Studies....................................................................................... 61
  4.1 Experiment 1: Advantage of multiple graphical representations ........................................... 62
  4.2 Experiment 2: Supporting representational sense-making processes .................................... 65
    4.2.1 Research questions and hypotheses .............................................................................. 65
    4.2.2 Methods ..................................................................................................................... 67
4.2.2.1. Experimental design ................................................................. 67
4.2.2.2. Participants .............................................................................. 69
4.2.2.3. Procedure ................................................................................ 69
4.2.2.4. Test instruments ................................................................. 69
4.2.3 Results ...................................................................................... 71
4.2.3.1. Analysis .................................................................................. 71
4.2.3.2. Effects of practice schedules ................................................. 73
4.2.4 Discussion ................................................................................ 76
4.3 Experiment 3: Supporting representational fluency-building processes ... 78
4.3.1 Classroom experiment: effects of practice schedules on learning outcomes ... 78
4.3.1.1. Research questions and hypotheses ........................................ 78
4.3.1.2. Methods ................................................................................ 79
  Experimental design ........................................................................... 79
  Participants ....................................................................................... 80
  Procedure ......................................................................................... 80
  Test instruments ............................................................................... 81
4.3.1.3. Results .................................................................................. 81
  Effects of practice schedules on students’ learning ................................ 81
  Effects of practice schedules on students’ benefit from multiple graphical representations 84
4.3.1.4. Summary ................................................................................ 85
4.3.2 Think-aloud study to investigate learning processes ...................... 86
4.3.3 Educational data mining to investigate learning processes ............. 88
  4.3.3.1. Bayesian knowledge tracing models ..................................... 90
  4.3.3.2. Evaluation results ................................................................. 92
  4.3.3.3. Summary ................................................................................ 93
4.3.4 Discussion ................................................................................ 94
4.4 Experiment 4: Supporting connectional sense-making and fluency-building processes ..... 97
4.4.1 Research questions and hypotheses .............................................. 99
4.4.2 Methods ................................................................................... 100
  4.4.2.1. Experimental design ............................................................... 100
  4.4.2.2. Participants .......................................................................... 100
  4.4.2.3. Procedure ............................................................................ 101
  4.4.2.4. Fractions Tutor versions ....................................................... 102
  4.4.2.5. Test instruments ................................................................. 103
4.4.3 Results ..................................................................................... 104
  4.4.3.1. Effects of connection-making support on learning outcomes .... 105
  Effects of sense-making support and fluency-building support ........... 106
  Effects of active engagement in sense-making processes .................. 107
  4.4.3.2. Causal path modeling ............................................................ 108
4.4.4 Discussion .................................................................................................................. 112

4.5 Experiment 5: Sequencing connectional sense-making and fluency-building processes... 114

4.5.1 Theoretical perspectives and hypotheses ................................................................. 114

4.5.1.1 Fluency-first hypothesis ....................................................................................... 116

4.5.1.2 Understanding-first hypothesis ............................................................................ 116

4.5.1.3 Specific predictions ............................................................................................. 117

Predictions on learning outcomes ..................................................................................... 117

Predictions on problem-solving behaviors ........................................................................ 119

Predictions on connection making and conceptual processing ......................................... 120

Predictions on visual attention .......................................................................................... 121

Predictions about relations between process-level and learning outcome measures ....... 123

4.5.2 Method ........................................................................................................................ 123

4.5.2.1 Experimental design ............................................................................................. 123

4.5.2.2 Participants .......................................................................................................... 124

4.5.2.3 Procedure ............................................................................................................. 124

4.5.2.4 Fractions Tutor versions ...................................................................................... 125

4.5.2.5 Test instruments .................................................................................................... 125

Learning outcome measures .............................................................................................. 125

Tutor log data .................................................................................................................... 126

Eye-tracking measures ...................................................................................................... 129

Retrospective interviews ................................................................................................... 131

4.5.3 Results ........................................................................................................................ 134

4.5.3.1 Effects on quiz: Connection-fluency and connection-understanding ...................... 137

4.5.3.2 Effects on posttest: Transfer of knowledge ............................................................. 139

4.5.3.3 Effects on learning curves: Rates of learning ......................................................... 140

4.5.3.4 Effects on retrospective interviews: Connection making and conceptual reasoning 143

4.5.3.5 Effects on eye-tracking measures: Visual attention ............................................ 144

4.5.3.6 Relations between dependent measures ............................................................... 147

Mediation of condition effects through performance on connection-fluency quiz .......... 148

Mediation of condition effects through problem-solving behaviors .............................. 151

4.5.4 Discussion .................................................................................................................. 153

5 Conclusion ................................................................................................................................... 157

5.1 Learning sciences perspective ....................................................................................... 158

5.1.1 Advantage of multiple graphical representations ....................................................... 158

5.1.2 Representational sense-making processes ................................................................. 160

5.1.3 Representational fluency-building processes ............................................................. 162

5.1.4 Connectional sense-making and fluency-building processes ................................... 164

5.1.5 Summary .................................................................................................................... 169

5.2 Educational technology perspective .............................................................................. 170
5.2.1 A successful intelligent tutoring system for fractions learning ................................. 170
5.2.2 Integrating learning sciences and intelligent tutoring systems research .................. 171
5.2.3 A principled methodology to resolving design conflicts ............................................ 173
5.3 Limitations and future research ......................................................................................... 175
5.4 Summary: An interdisciplinary multi-methods approach ................................................. 177
Acknowledgements .................................................................................................................... 178
Literature .................................................................................................................................. 180
Appendix 1: Pennsylvania State Standards ............................................................................. 212
Appendix 2: NCTM Standards ................................................................................................. 213
Appendix 3: Common Core Standards ................................................................................. 214
Appendix 4: Experiment 2 test items ....................................................................................... 216
Appendix 5: Experiment 3 test items ....................................................................................... 218
Appendix 6: Experiment 4 test items ....................................................................................... 220
Appendix 7: Experiment 5 test items ....................................................................................... 222
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), in part because different graphical 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 focused on learning with dual representations: 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 dual representations may in part be due to the belief that multiple representations are beneficial because they are presented in different symbol systems. 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 usu-
ally 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 focus on learning with dual representations, the educational psychology literature 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 visual information channel (Clark & Mayer, 2003; Leikin, Leikin, Waisman, & Shaul, 2013; Mayer, 2003; Mayer, 2005), which is known to 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 focused on dual representations, it is an open question how best to best design instructional materials with multiple graphical representations, and how to best to support students in using multiple graphical representations in a way that most benefits their robust learning.

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. Indeed, 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 educational technologies. 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 & Schnottz, 2009; Reimer & Moyer, 2005; Suh, Moyer, & Heo, 2005), and by providing individualized support on 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 which learning processes students need to engage in when 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 three categories. First, I provide a theoretical framework for learning with multiple graphical representations which extends prior frameworks on dual representations (see sections 2.2 and 5.1). My goal in providing a new theoretical framework is to explicitly describe the learning processes that students need to engage in when working with multiple representations using the same symbol system. Furthermore, the theoretical framework makes testable predictions which provide the foundation for my empirical work. In testing these predictions based on a series of controlled experiments (see section 4), I evaluate the theoretical framework. Finally, the theoretical framework has the potential to stimulate research on learning with multiple graphical representations in a variety of domains.

Second, I developed a successful educational technology: an intelligent tutoring system for fractions (see section 5.2). The Fractions Tutor (see section 3) uses multiple graphical represen-
tations 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). The development of the Fractions Tutor is based on a user-centered design methodology that integrates methods from learning sciences, intelligent tutoring systems, and human-computer interaction (see section 3.3). It incorporates the recommendations made by my theoretical framework for learning with multiple graphical representations and has been iteratively improved based on a sequence of controlled classroom experiments (see section 4).

Third, 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 based on my theoretical framework for learning with multiple graphical representations (see section 2.2) and have been empirically tested in a sequence of controlled experiments (see section 4). They address the gap between the focus on learning with dual representations in prior educational psychology research and the lack of guidance in common educational materials that include multiple graphical representations. Thereby, the instructional design principles provide guidance for the development of multi-representational educational technologies.

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, and a variety of quantitative and qualitative measures of learning processes, such as tutor log data, eye-tracking data, and think-aloud protocols. 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 the theoretical framework for my work (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 which is both the plat-
form 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 will 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.
1 Introduction

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 & Schnitz, 2009; Schnitz & 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; Schnitz & Bannert, 2003; Schwartz & Black, 1996; Zhang, 1997; Zhang & Norman, 1994) instructional materials tend to use not only one 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; Schnitz & Bannert, 2003).

By and large, the educational psychology literature on learning with multiple representations has taken a dual-representation approach: research has mostly focused on learning with textual descriptions and one additional graphical representation (Ainsworth & Loizou, 2003; Baetge & Seufert, 2010; Bodemer et al., 2005; Butcher & Aleven, 2008; Butcher, 2006; Eilam & Poyas, 2008; Eitel, Scheiter, & Schuler, 2013; Kuehl et al., 2010; Magnier et al., 2010; Mason, Pluchino, & Tornatora, 2013; Rasch & Schnitz, 2009; Suthers et al., 2008). Textual descriptions and graphical representations use different symbol systems (Schnitz & Bannert, 2003). The advantage of dual representations has been attributed to the fact that they stimulate deeper processing by requiring learners to integrate information across these different symbol systems. 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 dual 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 dual 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 dual 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 the cognitive tasks students need to accomplish when learning with dual representations are the same as those they engage in when learning with multiple graphical representations. Consequently, we do not know how best to support students in learning with multiple graphical representations, although 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 in is the unit, as opposed to the length of the entire number line (e.g., in a number line from 0 to 3). To compare the relative size of fractions using linear models, students have to judge the length of a linear segment relative to the defined unit of the representation. The goal in using 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 a rather abstract concept.

My research addresses these open questions about the processes involved in successful learning from multiple graphical representations and how to provide instructional support for these processes. Thereby, my research extends prior research on learning with dual representations to learning with multiple graphical representations, which is – as just argued – a common scenario in real-world instructional materials. Given that learning with dual representations from different symbol systems may involve different cognitive processes than learning with multiple graphical representations from the same symbol system, 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 developed a theoretical framework for learning with multiple graphical representations which extends existing theoretical frameworks on learning with dual representations (see section 2). This theoretical framework provides the foundation for the design of an educational technology for fractions learning (see section 3) and guides a sequence of experiments learning with multiple graphical representations of fractions (Rau, Aleven, & Rummel,
The goal of each experiment is (1) to investigate whether a hypothesized learning process is involved in successful learning with multiple graphical representations, (2) to investigate how best to design instructional support for this process, and (3) to develop and further refine the Fractions Tutor, an educational technology for fractions learning which incorporates these instructional design principles.
1.2 The 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 without 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 a technology 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 the 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.

In developing the Fractions Tutor, my goal was to develop a technology 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, superintendants). 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 Merrienboër, 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 research (Koedinger, 2002; Mayer, 2003; van Merrienboër 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 Merrienboër 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 Merrienboër et al., 2002).

Paying attention to such conflicting goals is a crucial prerequisite to developing an effective educational technology. If we fail to address stakeholders’ competing goals, students may dislike the technology 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 technology 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 technologies, dissemination and long-term success of the technology, 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 (Albert 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 educational technology design.

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 technologies.
1.3 Summary

In summary, my research integrates multiple perspectives which complement each other. First, my research evaluates a novel theoretical framework for learning with multiple graphical representations, and investigates how best to implement instructional support for processes students engage in when learning with 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 how to support learning with multiple graphical representations. Second, 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. To address these different goals, I have conducted several iterations of classroom experiments and lab studies with over 3,000 students. Each iteration served both to investigate theoretically motivated research questions about how best to support students’ learning through the use of multiple graphical representations and to iteratively improve the Fractions Tutor. 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) an empirically validated theoretical framework for learning with multiple graphical representations, (2) an effective educational technology for fractions learning, (3) a set of instructional design principles to guide the development of a range of multi-representational educational technologies, and (4) a new methodology for resolving design conflicts that often occur in real educational settings.
2 Theoretical Framework

To develop an educational technology that supports learning through adequate use of multiple graphical representations, it is crucial to first reflect on the cognitive tasks students need to accomplish in order to benefit from multiple representations. Simply integrating multiple graphical representations into an educational technology is unlikely to lead to the most optimal learning gains (Ainsworth, 2006; de Jong et al., 1998; Kim et al., 2013), but 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 focus on dual representations: they are based on research on learning with text and one additional graphical representation. None of them specifically focus on learning with multiple graphical representations. In this section, I first describe a number of existing frameworks for learning with dual representations, while explicitly pointing out their shortcomings and why their predictions may not apply to multiple graphical representations. 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.
2.1 *Existing theoretical frameworks for learning with dual representations*

When discussing theoretical models for learning with multiple representations, one needs to distinguish internal and external representations. External representations are “the knowledge and structure in the environment, as physical symbols, objects, or dimensions [...] embedded in physical configurations” (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 (ideally) they will then 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 & Schnott, 2009; Reed & Ettinger, 1987; Schnott & 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. Text is processed semantically through an analysis of its symbolic structure, leading to a propositional internal representation. Graphical representations, on the other hand, are processed perceptually, leading to a pictorial internal representation. During mental model formation, learners integrate both internal representations via structure mapping (Gentner & Markman, 1997). The integration process of information from different sign systems requires deep conceptual processing, which explains better learning from text and graphical representation than from text alone.

In line with the Dual Channel theory (Paivio, 1986), and building on Schnott and Bannert’s framework of text and graphic comprehension (Schnott & 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). Since the capacity of each information channel is limited but the capacity of both channels together is additive (Chandler & Sweller, 1991), learning with both text and graphical representations is more effective than learning with either text or graphical information alone.
representation makes better use of the learner’s mental capacity. Furthermore, active integration of the textual and picture-based internal representations into a coherent mental model requires deeper conceptual processing of the content, which leads to better learning.

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 du-
all representations (i.e., text and one graphical representation) lead to better learning, they do not predict an advantage of multiple graphical representations compared to a single graphical representation. Multiple graphical representations use the same sign system and employ the same organization principles. Schnitz 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 sign systems from which learners have to integrate information. Furthermore, since all graphical representations are processed in the visual information channel, multiple graphical representations does not increase a learner’s cognitive capacity. The Cognitive Theory of Multimedia Learning might even predict cognitive overload in the visual information channel 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 the same 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 that a circle shows 1/2, and a number line shows a segment from 0 to 2, with a dot at 1/2. A student might apply the part-whole approach to the number line and interpret the dot as showing 1/4 (i.e., taking the entire number line as the unit, rather than just the segment between 0 and 1). Knowing that the circle and the number line are both supposed to show the same fraction (i.e., 1/2) may help the student overcome the misconception that the entire segment shown by a number line (as opposed to the segment of just 0 to 1) denotes the unit. 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 explain advantages of multiple graphical representations over a single graphical representation, it does not specify the learning processes necessary to benefit from multiple graphical representations, 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 graphic (Bodemer & Faust, 2006; Bodemer et al., 2004; Plötzner, Bodemer, & Feuerlein, 2001; Plötzner et al., 2008; Schwonke et al., 2008; Seufert, 2003), or between symbolic representations and graphic (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 describe a new theoretical framework that specifies the learning processes students need to engage in so as to benefit from multiple graphical representations.
2.2 A new theoretical framework for 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 framework which describes processes involved in learning of robust domain knowledge. In almost any domain, learning involves sense-making processes. Sense-making processes are learning processes that lead to principled understanding of connections between multiple graphical representations based on their knowledge components. Knowledge components are units of knowledge that constitute the smallest divisible part of what a student can learn (much like an atom in physics). In fractions, the concepts of numerator, denominator, and the unit of a fraction (i.e., what the fraction is taken of) are examples of simple knowledge components. In order to effectively use multiple graphical representations to learn about a domain, 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 connec-
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). 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). 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), 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).

The left part of Fig. 3 illustrates these *representational sense-making processes* in developing *representational understanding* using fractions representations as an example. Here, students are presented with external graphical representations (e.g., a circle, a rectangle, and a number line). To develop understanding of each individual graphical representation, students need to relate each graphical representation to the domain-specific knowledge components it depicts (e.g., (Ainsworth, 2006; Noss et al., 1997). 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 colored sections, the denominator to the total number of sections, and the unit to the inherent shape of the representation. For the number line, the numerator is the number of sections between 0 and the dot, the denominator the number of sections between 0 and 1, and the unit is always the length between 0 and 1. During
Engaging in connectional sense-making processes means to establish relations between corresponding knowledge components of different graphical representations. For example, a student might reason that a circle with one colored section shows the numerator of 1, and that a number line with one section between 0 and the dot shows a numerator of 1; that a circle with two total sections shows a denominator of 2, that the number line with two sections between 0 and 1 shows a denominator of 2; and that, since both graphical representations show the same numerator and the same denominator, they both show 1/2. The student reasons on the basis of the connections between the graphical representations at the level of the knowledge components numerator and denominator. The right part of Fig. 3 illustrates connectional sense-making processes. The ability to make connections between multiple representations are 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).

However, learning does not only rely on understanding: knowledge is only useful if it is readily accessible whenever needed. A learner who has readily accessible knowledge is said to have fluency in that knowledge (Koedinger et al., 2012). Typically, fluency is considered as the ability to retrieve facts from memory (Arroyo, Royer, & Woolf, 2011). Perceptual fluency, on the other hand, 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). 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). 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. Kellman and colleagues (Kellman et al., 2009) describe the importance of fluency in using individual graphical representations (henceforth *representational fluency-building processes* leading to *representational fluency*), and fluency in making connections between different graphical representations (henceforth *connectional fluency-building processes* leading to *connectional fluency*).

Accordingly, *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). 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.

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). Fluent learners can treat one graphical representation as a single perceptual chunk, which allows them to perform quickly and effortlessly with multiple graphical representations. Hence, 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.

The theoretical framework just described serves as a basis for a sequence of experimental studies (see section 4), each of which is carried out within the context of the Fractions Tutor. It also crucially informed the instructional design of the Fractions Tutor (see section 3), which serves as the research platform for each of the experimental studies.
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 conceptual 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 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 both counting (Fuson, 1988) and 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), 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. 4). 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. At the same time, it does not have a standard shape for the unit, like the circle does. 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. 4. Interactive circle, rectangle, and number line representations, as used in the Fractions Tutor.](image)

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 (described 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 (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, Svhila, & 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.

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 User-centered design

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, they do inevitably arise in the complex context of real educational settings. And 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.

Before describing this approach to resolve conflicts between stakeholder goals in the Fractions Tutor, I will 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.

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). Interviews with teachers confirmed the need for an effective educational technology that can 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 technology). These assessment tasks guided the selection of practice problems (i.e., problems students should solve as part of the educational technology). 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.

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 Merrienboër, Yates, & Early, 2007; Koedinger et al., 2012; Schraagen, Chipman, & Shalin, 2000). Cognitive task analysis seeks to identify the knowledge components (i.e., units of knowledge) 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. To develop the Fractions Tutor, I used Cognitive Tutor Authoring Tools (CTAT, Aleven, 2010; Aleven, McLaren, Sewall, & Koedinger, 2006; Aleven, McLaren, Sewall, & Koedinger, 2008), which allows for rapid prototyping and fast implementation of iterative design changes. I included teachers in the development process. Involving teachers not only improved the quality of the educational technology – it also helped establish relations with teachers and revealed further stakeholder goals.

The outcome of stage 4 is a set of working Cognitive Tutor problems ready for further testing.

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. A set of methods are available 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 technology in classrooms is indispensable. 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.
tions Tutor provided a useful basis for identifying issues in usability and difficulty level of particular steps within the educational technology.

As part of this stage, I conducted a series of parametric studies in classrooms (described in detail in section 4), which investigated instructional support for the learning processes proposed by my theoretical framework for learning with multiple graphical representations (described in section 2.2). As a consequence of these studies, the Fractions Tutor uses graphical representations in the following manner:

- The Fractions Tutor provides support representational understanding by encouraging students to relate graphical representations (e.g., circles) to symbolic representations (e.g., ½) in the form of menu-based prompts (see Experiment 1, section 4.1, Rau et al., 2009).
- The Fractions Tutor supports representational understanding through the use of 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 supports representational fluency by frequently switching between different graphical representations (see Experiment 3, section 4.3, Rau, Rummel et al., 2012).
- The Fractions Tutor provides support for connectional understanding by encouraging students to relating 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 combines support for connectional understanding, provided before support connectional fluency (see Experiments 4 and 5, sections 4.4 and 4.5, Rau, Aleven, & Rummel, 2013a; Rau, Aleven et al., 2012).

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.

3.3.6. Stage 6: Classroom use and evaluation

The goal of stage 6 is to evaluate the educational technology in the field. Randomized field trials are the method of choice 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 ran-
randomly selected teachers and student or teacher surveys further helped identifying problem-solving behaviors and learning processes. In addition, the analysis of student 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. Finally, I present three approaches to resolving these 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” (Catrambone & Holyoak, 1998; Singley, 1990). The Cognitive Theory of Multimedia Learning (Mayer, 2003; Mayer, 2005) suggests to use color only sparingly, and to highlight only conceptually relevant aspects of the problem (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 technologies 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 technology difficult to use for students results in teachers helping students out with usability issues rather than helping with the content. An educational technology 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. I then 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.

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

Table 1. Overview of example goals, instructional design principles, and 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, before the background of students’ poor reading ability, 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 abilities. Teachers were busy helping students understand problem statements, rather than helping them with the math problems.

Finally, conflicts 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 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 consider again 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 our 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. 5 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 hue, 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. 5 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, $\chi^2 (2, N = 429) = 236.86, p < .001.$

**Fig. 5.** 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 (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 which 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). I describe this experiment in detail below (see Experiment 4, section 4.4).

3.3.3.3. Cross-iteration studies

Unfortunately, it is not possible to conduct a controlled experiment for every design decision. In this case, I recommend conducting cross-iteration studies. 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. 6 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. 7 shows an example of the next iteration of the Fractions Tutor, without cover stories.

![Fractions Tutor Example](image)

**Fig. 6.** Example of an early version of the Fractions Tutor with cover stories. Students learn about fractions in the context of sharing pizzas.

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. 8). 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 a lot 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.
Fig. 8. Introductory problems with and without cover stories to introduce how the number line depicts fractions.
### 3.4 Overview of curriculum

**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
- **NCTM Standards Grades 3-5:** #1, #3
- **Common Core Standard:** 3.NF.1, 3.NF.2, 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
- **NCTM Standards Grades 3-5:** #1, #3
- **Common Core Standard:** 3.NF.1, 3.NF.2, 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
- **NCTM Standards Grades 3-5:** #1, #3
- **Common Core Standard:** 3.NF.1, 3.NF.2, 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, #3
- **Common Core Standard:** 3.NF.1, 3.NF.2, 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
- **NCTM Standards Grades 3-5:** #1, #3
- **Common Core Standard:** 3.NF.1, 3.NF.2, 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
- **NCTM Standards Grades 3-5:** #1, #3
- **Common Core Standard:** 3.NF.1, 3.NF.2

---

1. see Appendix 1 for Pennsylvania State Standard definitions
2. see Appendix 2 for NCTM Standard definitions
3. see Appendix 3 for Common Core Standard definitions
<table>
<thead>
<tr>
<th>Topic 3: Reconstructing The Unit</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem type:</strong> Given a unit fraction, reconstruct the unit of the fraction (using circle diagrams, rectangles, and number lines)</td>
<td>Common Core Standard 3.NF.3</td>
</tr>
</tbody>
</table>
| **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 |

<table>
<thead>
<tr>
<th>Topic 4: Naming Improper Fractions</th>
<th></th>
</tr>
</thead>
</table>
| **Problem type:** Determine what improper fraction a given graphical representation shows (using circle diagrams, rectangles, and number lines) | 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 |
| **Learning goal:** Understand that fractions can be larger than 1 |  |
| **Problem type:** Determine what fraction a given graphical representation represents, given the unit of the fraction (using circle diagrams, rectangles, and number lines) | 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 |
| **Learning goal:** Understand that the unit determines the relative size of the fraction |  |

<table>
<thead>
<tr>
<th>Topic 5: Making Graphical Representations of Improper Fractions</th>
<th></th>
</tr>
</thead>
</table>
| **Problem type:** Construct a graphical representation for a fraction given symbolically (using circle diagrams, rectangles, and number lines) | 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 |
| **Learning goal:** Become familiar with the (interactive) graphical representations and link it to the symbolic representation |  |
| **Problem type:** Compare fractions given a graphical representation and the corresponding symbolic fraction (using circle diagrams, rectangles, and number lines) | 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 |
| **Learning goal:** Developing a sense for the size of fractions using the graphical representation and link it to the symbolic representation |  |

<table>
<thead>
<tr>
<th>Topic 6: Equivalent Fractions: Underlying Concepts</th>
<th></th>
</tr>
</thead>
</table>
| **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 | 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 |
| **Learning goal:** Understand the invariance of amounts when partitioning a fraction into more sections |  |
| **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) | 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 |
| **Learning goal:** Understand that multiplying numerator and denominator by 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

<table>
<thead>
<tr>
<th>Problem type:</th>
<th>Given a graphical representation of a proper fraction (circle diagrams, rectangles, and number lines), manipulate it to expand it and reduce it</th>
<th>Learning goal:</th>
<th>Understand that expanding and reducing fractions are interchangeable activities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PA State Standard 2.1.8.A</td>
<td></td>
<td>NCTM Standards Grades 3-5 #1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Common Core Standard 3.NF.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Common Core Standard 4.NF.1</td>
</tr>
</tbody>
</table>

### Topic 8: Comparing Fractions

**Problem type:** Given two fractions and their graphical representation (circle diagrams, rectangles, or number lines), use common benchmarks (e.g., $\frac{1}{2}$, $\frac{1}{4}$) to compare them  
**Learning goal:** Being able to use benchmarks to compare fractions

<table>
<thead>
<tr>
<th>Problem type:</th>
<th>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</th>
<th>Learning goal:</th>
<th>Being able to use equivalent fractions to compare fractions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PA State Standard 2.2.8.B</td>
<td></td>
<td>NCTM Standards Grades 3-5 #1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NCTM Standards Grades 3-5 #2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Common Core Standard 3.NF.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Common Core Standard 4.NF.2</td>
</tr>
</tbody>
</table>

### 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

<table>
<thead>
<tr>
<th>Problem type:</th>
<th>Add two given fractions with the different denominators and specify the unit of the addend fractions and the sum fraction</th>
<th>Learning goal:</th>
<th>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</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PA State Standard 2.2.8.B</td>
<td></td>
<td>Common Core Standard 5.NF.1</td>
</tr>
</tbody>
</table>

### 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

<table>
<thead>
<tr>
<th>Problem type:</th>
<th>Subtract two given fractions with the different denominators and specify the unit of the subtrahend fractions and the difference fraction</th>
<th>Learning goal:</th>
<th>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 the common denominator of two subtrahend fractions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PA State Standard 2.2.8.B</td>
<td></td>
<td>Common Core Standard 5.NF.1</td>
</tr>
</tbody>
</table>

| **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. The activities and concepts covered in each topic align with U.S. education standards including the NCTM standards and common core standards (see Table 2). 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 part of 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, which a teacher present to help out individual students who need help. The Fractions Tutor is available for free on a website for middle school math (Aleven, McLaren, & Sewall, 2009). 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 an 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. 9 shows a tutor problem which 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. 9), by highlighting, and by dragging-and-dropping sections (see Fig. 4). The Fractions Tutor employs subgoaling (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. 9 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. 9).

![Fig. 9. Naming the fraction shown in an interactive circle, with reflection prompts.](image)

3.4.2 Connectional sense-making support for connectional understanding

Fig. 10 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. 10). 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 Merrienboer, 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 dual 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 Merrienboer, 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).

To ensure that students read through the worked examples, they are asked to fill in the last step themselves (step A-3 in Fig. 10). Once they complete that step, the problem-solving part of the worked example appears on the right (part B in Fig. 10). 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.

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. 10). 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. 10. Connectional sense-making support problem for equivalent fractions.

Fig. 11. Connectional sense-making support problem for fraction comparison.

Fig. 11 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 that support described for the equivalent fractions example (Fig. 10).

**3.4.3 Connectional fluency-building support for connectional fluency**

![Fig. 12. Connectional fluency-building support problem for equivalent fractions.](image)

![Fig. 13. Connectional fluency-building support problem for fraction comparison.](image)

Fig. 12 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. 13 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 based on 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. 14 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. 15 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 which include transfer items. It is usable within real classroom settings and addresses the goals and needs of both students and teachers.
Fig. 14. Learning gains on the conceptual knowledge test.

Fig. 15. 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) whether the processes described as part of the theoretical framework (see section 2) play a role in students’ robust learning of fractions, and (2) how best to support these processes within an intelligent tutoring system. At the same time, (3) each experiment serves to iteratively improve the Fractions Tutor (see section 3).

Taken together, each classroom experiment provides evidence for a particular aspect of the theoretical framework. Furthermore, each experiment leads 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 instructional support for processes involved in learning with multiple graphical representations. Table 3 gives an overview of the research hypotheses (as deduced from the theoretical framework) addressed in each experiment, of the instructional design principles that follow from the experiments, and how these principles are implemented in the Fractions Tutor.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Principle</th>
<th>Implementation</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple graphical representations lead to better learning than a single graphical representation</td>
<td>Use multiple graphical representations to support robust learning</td>
<td>Circles, rectangles, number lines</td>
<td>Experiments 1 &amp; 3</td>
</tr>
<tr>
<td>To benefit from multiple graphical representations, students need to be supported in representational sense-making processes</td>
<td>Use reflection prompts to support representational understanding</td>
<td>Menu-based reflection prompts</td>
<td>Experiment 1</td>
</tr>
<tr>
<td></td>
<td>Interleave task types to support representational understanding</td>
<td>Frequently switch between different task types</td>
<td>Experiment 2</td>
</tr>
<tr>
<td>To benefit from multiple graphical representations, students need to be supported in representational fluency-building processes</td>
<td>Interleave representations to support representational fluency</td>
<td>Frequently switch between graphical representations</td>
<td>Experiment 3</td>
</tr>
<tr>
<td>To benefit from multiple graphical representations, students need to be supported in connectional sense-making processes and in connectional fluency-building processes</td>
<td>Combine worked examples and mixed representations to support connectional understanding and connectional fluency</td>
<td>Worked examples and mixed representations problems</td>
<td>Experiment 4</td>
</tr>
<tr>
<td></td>
<td>Provide worked-example support for connectional understanding before mixed representations support for connectional fluency</td>
<td>Worked examples and mixed representations problems</td>
<td>Experiment 5</td>
</tr>
</tbody>
</table>

Table 3. Instructional design principles and their implementation in the Fractions Tutor.
4.1 Experiment 1: Advantage of multiple graphical representations

As discussed in section 1.1, research shows that dual representations can significantly enhance students’ learning: students typically learn better from a combination of text and graphics than from text alone (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 dual representations is mediated by an increased engagement in self-explanation activities (i.e., the process of generating explanations to oneself with the goal to make sense of what one is learning (Chi, Bassok, Lewis, Reimann, & Glaser, 1989): students who generate more high-quality self-explanations also show the highest learning gains when working with dual representations (Ainsworth & Loizou, 2003). Based on these findings, Ainsworth and Loizou (2003) hypothesize that dual 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 dual 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 dual 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 reflection prompts, or whether the advantage of multiple representations depends on students receiving reflection prompts.

Furthermore, all these prior studies were conducted on learning with dual representations: 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 general-
ize 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 addresses these questions and thereby lays the foundation for my dissertation research. Specifically, Experiment 1 systematically investigates whether the advantage of dual representations generalizes the more complex, multi-representational learning materials that are commonly used in real educational settings: multiple graphical representations. Furthermore, Experiment 1 tests whether students’ benefit from multiple graphical representations depends on receiving instructional support for representational sense-making processes by the means of prompts to 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 help students relate the graphical representation to the symbolic notation while emphasizing knowledge components such as numerator and denominator.

Results based on pretest, immediate and delayed posttest scores from 112 students show no main effect of number of graphical representations, but a main effect of self-explanation prompts on reproduction of conceptual knowledge 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 with multiple graphical representations, whereas students within the no-prompt conditions performed worse when provided with multiple graphical representations, compared to learning with a single graphical representation. Please refer to Rau and colleagues (2009) or Rau (2008) for a more detailed description of Experiment 1.

Experiment 1 was part of my diploma thesis (the German equivalent to a Master thesis), see (Rau, 2008).
In summary, Experiment 1 demonstrates that representational understanding is an important prerequisite for the effectiveness of multiple graphical representations. Reflection prompts are a successful means to support representational sense-making processes by helping students relate graphical representations to symbolic representations at the level of corresponding knowledge components. This type of instructional support for representational sense-making processes enables students to benefit from multiple graphical representations in acquiring robust knowledge about fractions.
4.2 Experiment 2: Supporting representational sense-making processes

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?

This decision is likely to impact students’ representational sense-making processes. As I describe in the following, the literature on practice schedules suggests that the sequence in which graphical representations and task types are provided to students affects which aspects of the learning material students process more deeply, and hence, which aspects of the learning material are the object of their sense-making processes. In other words, the practical decision of how to sequence graphical representations and task types is expected to influence students' representational understanding.

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’ learning. Generally, contextual interference research demonstrates that “interleaved practice” leads to better learning results 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), 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 Merrienboër, & Paas, 1998; van Merriënboer, Schuurman, de
Croock, & Paas, 2002) and decision-making tasks (Helsdingen, van Gog, & van Merrienboër, 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 are interleaved, or when task types are interleaved.

In particular, the advantage of interleaved practice has been attributed to processes which 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 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).

When multiple graphical representations are presented across different task types, the sequence of graphical representations and task types will affect which aspects of the learning material students will process deeply. I hypothesize that blocking representations while interleaving task types will allow students to see more clearly how each individual graphical representation represents fractions across a variety of different task types, allowing them to engage in representational sense-making processes. Therefore, if students’ representational understanding plays an important role in their robust learning of domain knowledge from multiple graphical representations, I expect the best learning gains when students learn with a version of the Fractions Tutor
that blocks graphical representations while interleaving task types, as opposed to a version that blocks task types while interleaving graphical representations.

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
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 intelligent tutoring system. 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. 16 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. 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. 16), 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

5 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.
encountered a new task type after every single problem. By contrast, in the int-reps condition (see the table on the right in Fig. 16), 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.

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. 16 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 the Fractions Tutor (with the version depicted in Fig. 6, see section 3.3.3.3, on topics 1, 2, and 6-8 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 4 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 fraction 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, analogical 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. Finally, the tests included items we 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 confirmed that both test versions were equally difficult. 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 effectiveness and efficiency measures. The effectiveness 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 using the following formula:

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

(1)

70
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 effectiveness 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.

<table>
<thead>
<tr>
<th></th>
<th>pretest</th>
<th>immediate posttest</th>
<th>delayed posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Representational</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>effectiveness int-types</td>
<td>.55 (.20)</td>
<td>.61 (.27)</td>
<td>.59 (.24)</td>
</tr>
<tr>
<td></td>
<td>.53 (.22)</td>
<td>.52 (.26)</td>
<td>.39 (.32)</td>
</tr>
<tr>
<td>operational effectiveness int-types</td>
<td>.38 (.31)</td>
<td>.51 (.34)</td>
<td>.44 (.36)</td>
</tr>
<tr>
<td></td>
<td>.43 (.33)</td>
<td>.40 (.35)</td>
<td>.39 (.30)</td>
</tr>
<tr>
<td>operational time-on-task int-types</td>
<td>100.18 (29.53)</td>
<td>70.89 (22.46)</td>
<td>66.95 (23.60)</td>
</tr>
<tr>
<td></td>
<td>103.63 (32.19)</td>
<td>74.88 (28.33)</td>
<td>79.24 (30.52)</td>
</tr>
<tr>
<td><strong>Operational</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>effectiveness int-types</td>
<td>-.35 (.89)</td>
<td>.48 (.83)</td>
<td>.52 (.80)</td>
</tr>
<tr>
<td></td>
<td>-.48 (.98)</td>
<td>.16 (.90)</td>
<td>-.31 (1.01)</td>
</tr>
<tr>
<td>operational time-on-task int-types</td>
<td>-.50 (.90)</td>
<td>.31 (.83)</td>
<td>.24 (.77)</td>
</tr>
<tr>
<td></td>
<td>-.42 (1.12)</td>
<td>.12 (.78)</td>
<td>.17 (.63)</td>
</tr>
</tbody>
</table>

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

### 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 we did not have information on how much time they spent on each item of the test (n = 5). 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, $\chi^2 (1, N = 158) < 1$, 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 (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 differences between classes, and at level 4, I accounted for differences between schools. In addition, I used post-hoc comparisons to clarify the effect of blocking versus interleaving. More specifically, the following hierarchical linear model was fitted to the data:

$$Y_{ijkl} = (((\mu_{0000} + W_{0001}) + V_{00kl}) + \beta_2 * c_{0jkl} + U_{0jkl}) + \beta_1 * t_{ijkl} + \beta_2 * c_{0jkl} * t_{ijkl} + R_{ijkl}$$

with

(level 1) $Y_{ijkl} = \beta_{0ijkl} + \beta_1 * t_{ijkl} + \beta_2 * c_{0ijkl} * t_{ijkl} + R_{ijkl}$

(level 2) $\beta_{0ijkl} = \delta_{00kl} + p_{0ijkl} + c_{0ijkl} + \beta_3 * c_{0ijkl} + + U_{0ijkl}$

72
where the dependent variable $Y_{ijkl}$ is student $i$’s score on the dependent measures at test $t_i$ (i.e., immediate or delayed posttest), $\beta_{0jkl}$ is student $j$’s score across tests, $\beta_1$ is the effect of test time, $\beta_2$ being the effect of the interaction of condition with test time, $\delta_{00kl}$ is the average performance of class $k$, $p_{0jkl}$ is student $j$’s performance on the pretest, $\epsilon_{0jkl}$ indicates how many problems a student solved on the relevant knowledge components being tested by either the representational knowledge test or the operational knowledge test, $\beta_3$ being the effect of condition, $\beta_4$ is the effect of the interaction between student $j$’s performance on a pretest, $\gamma_{000l}$ is the average performance of school $l$, and $\mu_{0000}$ is the overall average.

The reported $p$-values were adjusted using the Bonferroni correction. I report partial $\eta^2$ 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 $\eta^2$ 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
To investigate the effect of practice schedules on effectiveness of representational knowledge, I applied the HLM in equation 2 to students’ effectiveness scores on the representational knowledge subscale of the test. Table 5 provides an overview of the learning results on both the representational effectiveness and representational efficiency measures. Table 6 summarizes the least squared means and standard deviations generated by the model for representational effectiveness and representational efficiency. There was a significant main effect for condition on representational effectiveness, $F(1, 100) = 18.28, p < .01$, 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 $\eta^2 < .03$. Post-hoc comparisons between groups were computed to clarify the interaction effect at the immediate posttest and the delayed posttest, respectively. On representational effectiveness, 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 effectiveness of representational knowledge.

<table>
<thead>
<tr>
<th>measure</th>
<th>test-time</th>
<th>main effects / interaction effects</th>
<th>tendency of pairwise comparisons</th>
<th>significant (yes/no)</th>
<th>F/t-value</th>
<th>adj. p-value</th>
<th>effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>condition</td>
<td>yes</td>
<td>$F(1, 100) = 23.97$</td>
<td>p &lt; .01</td>
<td>partial $\eta^2 = .11$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>test-time</td>
<td>yes</td>
<td>$F(1, 100) = 4.99$</td>
<td>p &lt; .05</td>
<td>partial $\eta^2 = .01$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>condition * test-time</td>
<td>yes</td>
<td>$F(1, 100) = 7.32$</td>
<td>p &lt; .01</td>
<td>partial $\eta^2 = .05$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>immediate posttest</td>
<td>int-types &gt; int-reps</td>
<td>yes</td>
<td>$t(100) = 2.34$</td>
<td>p &lt; .05</td>
<td>$d = .37$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>delayed posttest</td>
<td>int-types &gt; int-reps</td>
<td>yes</td>
<td>$t(100) = 5.55$</td>
<td>p &lt; .01</td>
<td>$d = .88$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| condition | yes | $F(1, 100) = 18.28$ | p < .01 | partial $\eta^2 = .07$ |
| test-time | yes | $F(1, 100) = 4.94$ | p < .05 | partial $\eta^2 = .02$ |
| condition * test-time | yes | $F(1, 100) = 7.46$ | p < .01 | partial $\eta^2 = .03$ |
| 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 5.** Overview of study results on differences between conditions on representational effectiveness and representational efficiency.

<table>
<thead>
<tr>
<th></th>
<th>immediate posttest</th>
<th>delayed posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>representational effectiveness</td>
<td>int-types</td>
<td>.58 (.05)</td>
</tr>
<tr>
<td></td>
<td>int-reps</td>
<td>.50 (.05)</td>
</tr>
<tr>
<td>representational efficiency</td>
<td>int-types</td>
<td>.41 (.22)</td>
</tr>
<tr>
<td></td>
<td>int-reps</td>
<td>.03 (.22)</td>
</tr>
</tbody>
</table>

**Table 6.** Least squared means and standard deviations (in parentheses) for representational effectiveness and representational efficiency at immediate posttest, delayed posttest by condition.

To investigate the effect of practice schedules on effectiveness 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 $\eta^2 = .05$. Post-hoc 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 effectiveness of representational knowledge.

<table>
<thead>
<tr>
<th>test scale</th>
<th>test-time</th>
<th>main effects / interaction effects</th>
<th>tendency of pairwise comparisons</th>
<th>significant (yes/no)</th>
<th>( F/t)-value</th>
<th>adj. ( p)-value</th>
<th>effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>operational effectiveness</td>
<td>condition</td>
<td>no</td>
<td>( F(1,100) = 2.05 )</td>
<td>( p &gt; .10 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>test-time</td>
<td>yes</td>
<td>( F(1,100) = 3.04 )</td>
<td>( p &lt; .10 )</td>
<td>partial ( \eta^2 &lt; .01 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>condition * test-time</td>
<td>no</td>
<td>( F(1,100) = 1.36 )</td>
<td>( p &gt; .10 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>immediate posttest</td>
<td>int-types &gt; int-reps</td>
<td>no</td>
<td>( t &lt; 1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>delayed posttest</td>
<td>int-types &gt; int-reps</td>
<td>no</td>
<td>( t &lt; 1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>operational efficiency</td>
<td>condition</td>
<td>no</td>
<td>( F &lt; 1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>test-time</td>
<td>no</td>
<td>( F &lt; 1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>condition * test-time</td>
<td>no</td>
<td>( F &lt; 1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>immediate posttest</td>
<td>int-types &gt; int-reps</td>
<td>no</td>
<td>( t &lt; 1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>delayed posttest</td>
<td>int-types &gt; int-reps</td>
<td>no</td>
<td>( t &lt; 1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Overview of study results on differences between conditions, obtained from HLM described in equation 2.

<table>
<thead>
<tr>
<th>operational effectiveness</th>
<th>immediate posttest</th>
<th>delayed posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>int-types</td>
<td>.51 (.04)</td>
<td>.43 (.04)</td>
</tr>
<tr>
<td>int-reps</td>
<td>.41 (.04)</td>
<td>.39 (.04)</td>
</tr>
</tbody>
</table>

Table 8. Least squared means and standard deviations (in parentheses) for operational effectiveness and operational efficiency at immediate posttest, delayed posttest by condition.

To investigate the effects of practice schedules on effectiveness of operational knowledge, I applied the HLM in equation 2 to students’ effectiveness scores on the operational knowledge subscale of the test. Table 7 provides an overview of the learning results on both the operational effectiveness and operational efficiency measures. Table 8 summarizes the least squared means and standard deviations generated by the model for operational effectiveness and operational efficiency. On operational effectiveness, 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 effectiveness,
$F(1,100) = 3.04, p < .10$, partial $\eta^2 < .01$. There was no significant interaction effect for operational effectiveness, $F(1,100) = 1.36, p > .10$. These results show that the int-types condition does not outperform the int-reps condition on effectiveness 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 sub-scale 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.

4.2.4 Discussion

Taken together, Experiment 2 shows that interleaving task types while blocking graphical representations leads to better effectiveness and efficiency of representational knowledge than interleaving graphical representations while blocking task types. How might one explain the differences between conditions on representational knowledge? As argued, interleaved practice increases the need for repeated reactivation of knowledge, strengthens that knowledge and increases the likelihood of long-term retention. Furthermore, interleaved practice increases the number of opportunities for abstraction. Interleaving task types may encourage students to abstract across different applications of the same graphical representations. 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 robust representational knowledge than interleaving representations. Taken together, Experiment 2 shows that the reactivation and abstraction across task types is more effective at enhancing representational sense-making processes that lead to representational understanding.

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 stu-
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 indicate that practice schedules have a greater impact on conceptual knowledge than on operational knowledge.

In conclusion, the results from Experiment 2 are of both theoretical and practical importance. At the theoretical level, Experiment 2 provides evidence that students’ *representational understanding* is an important aspect of robust conceptual learning of domain knowledge. At the practical level, Experiment 2 provides guidance for instructional designers for how to implement support for this learning process: interleaving task types while blocking representations is an effective means to support representational sense-making processes. This finding extends the literature on learning with multiple representations by showing that practice schedules of multiple graphical representations have an effect on students’ learning. This finding also extends prior research on practice schedules by showing that the choice of aspect to interleave impacts what is learned.
4.3 Experiment 3: Supporting representational fluency-building processes

While Experiment 2 investigates the effects of practice schedules on representational sense-making processes, Experiment 3 investigates the effects of practice schedules on representational fluency-building processes. Experiment 2 shows that interleaving task types while blocking graphical representations promotes students' robust learning by enhancing representational understanding. It remains open whether interleaving graphical representations, in addition to interleaving task types, can further promote students’ learning of domain knowledge.

In this section, I first describe the results based on the learning outcomes which show that interleaving graphical representations (in addition to interleaving task types) enhance students' robust learning of fractions. I will then describe additional findings obtained from the analysis of process-level measures which demonstrate that the advantage of interleaving graphical representations is due to effects on representational fluency-building processes.

4.3.1 Classroom experiment: effects of practice schedules on learning outcomes

Experiment 3 addresses the question: does interleaving graphical representations (in addition to interleaving task types) enhance students' robust learning by promoting representational fluency?

4.3.1.1. Research questions and hypotheses

As argued in Experiment 2, blocked practice with multiple graphical representations promotes representational understanding as students can get in-depth experience in using one graphical representation across a sequence of (interleaved) task types. Interleaving multiple graphical representations, on the other hand, should allow students to frequently reactivate their knowledge about each graphical representation, which will strengthen their memory and increase the likelihood that students can fast and effortlessly use their knowledge about individual graphical representations in subsequent learning tasks. In other words, interleaved practice is expected to increase students’ representational fluency. Furthermore, if representational fluency plays an important role in students' learning, then students' benefit from multiple graphical representations (compared to a single graphical representation) should depend on receiving the optimal practice schedule.

Specifically, Experiment 3 investigates the following hypotheses:
Hypothesis 1: Interleaved practice schedules of multiple graphical representations enhance students’ learning of robust knowledge of fractions.

Hypothesis 2: Interleaved practice schedules practice schedules of multiple graphical representations enhance students’ benefit from multiple graphical representations compared to 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.

Experimental design

I contrasted four practice schedules of multiple graphical representations, and three single-graphical representations control conditions. Fig. 17 illustrates the practice schedules of task

Fig. 17. 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.
types and graphical representations for the four multiple graphical representations conditions. 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. Following the results from Experiment 2, each task type was visited three times. I randomly assigned students to one of seven 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 the order in which students encountered the graphical representations. Finally, students in the three single graphical representation conditions worked 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. I had to apply this stringent criterion to ensure that students in the blocked condition encountered all three graphical representations (see Fig. 17). This results in a total of \( N = 290 \) (\( n = 63 \) in blocked, \( n = 53 \) in moderate, \( n = 52 \) in fully interleaved, \( n = 62 \) in increased, \( n = 21 \) in single-circle, \( n = 20 \) in single-rectangle, \( n = 19 \) in single-number-line).

Procedure
Experiment 3 took place at the end of the school year 2009/2010. 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. 7, 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 post-test 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: fluency with area models (i.e., circles and rectangles), fluency 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. Conceptual transfer items included proportional reasoning questions with and without graphical representations. Procedural transfer items included comparison questions with and without graphical representations. 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 5.

**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, \( ps > .10 \). There was no significant effect for order of multiple graphical representations for any dependent measure, \( F(5, 285) = 1.56, ps > .10 \).

I used a hierarchical linear model (HLM, see Raudenbush & Bryk, 2002) with four nested levels to analyze the data in order to take into account nested sources of variance (due to the fact that a student’s performance can be partially explained by his/her class and school. I modeled performance on the tests for each student (level 1), differences between students nested within
classes (level 2), differences between classes nested within schools (level 3), and differences between schools (level 4). More specifically, the following HLM model was fitted to the data:

\[ Y_{ijkl} = (((μ_{0000} + W_{000l} + V_{00kl}) + β_2 * c_{oijkl} + U_{oijkl}) + β_1 * t_{ijkl} + β_2 * c_{oijkl} * t_{ijkl} + R_{ijkl} ) \]

with

(level 1) \[ Y_{ijkl} = β_{oijkl} + β_1 * t_{ijkl} + β_2 * c_{oijkl} * t_{ijkl} + R_{ijkl} \]

(level 2) \[ β_{oijkl} = δ_{00kl} + p_{oijkl} + β_3 * c_{oijkl} + β_4 * p_{oijkl} * c_{oijkl} + U_{oijkl} \]

(level 3) \[ δ_{00kl} = γ_{000l} + V_{00kl} \]

(level 4) \[ γ_{000l} = μ_{0000} + W_{000l} \]

where the dependent variable \( Y_{ijkl} \) is student \( i \)’s score on the dependent measures at test \( t_i \) (i.e., immediate or delayed posttest), \( β_{oijkl} \) is student \( i \)’s score across tests, \( β_1 \) is the effect of test time, \( β_2 \) being the effect of the interaction of condition with test time, \( δ_{00kl} \) is the average performance of class \( k \), \( p_{oijkl} \) is student \( j \)’s performance on the pretest, \( β_3 \) being the effect of condition, \( γ_{000l} \) is the average performance of school \( l \), and \( μ_{0000} \) is the overall average. I included the interaction of students’ pretest scores with condition \( (β_4 * p_{oijkl} * c_{oijkl}) \) to investigate aptitude-treatment interaction effects (i.e., whether students’ benefits from one condition over the other depends on their prior knowledge).

Since the HLM described in (2) 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 included pretest score in the dependent variable, yielding:

\[ Y_{ijkl} = (((μ_{0000} + W_{000l} + V_{00kl}) + β_2 * c_{oijkl} + U_{oijkl}) + β_1 * t_{ijkl} + β_2 * c_{oijkl} * t_{ijkl} + R_{ijkl} ) \]

with

(level 1) \[ Y_{ijkl} = β_{oijkl} + β_1 * t_{ijkl} + β_2 * c_{oijkl} * t_{ijkl} + R_{ijkl} \]

(level 2) \[ β_{oijkl} = δ_{00kl} + β_3 * c_{oijkl} + U_{oijkl} \]

(level 3) \[ δ_{00kl} = γ_{000l} + V_{00kl} \]

(level 4) \[ γ_{000l} = μ_{0000} + W_{000l} \]

where the dependent variable \( Y_{ijkl} \) is student \( j \)’s score on the dependent measures at test \( i \) (i.e., pretest, immediate posttest, or delayed posttest).

I used planned contrasts and post-hoc comparisons 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
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 understanding of area models, $F(3, 867) = 2.57, p < .10$, $\eta^2 = .01$, 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.

To clarify the interaction between test time and condition, I used post-hoc 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 against the remaining three multiple graphical representations conditions taken together, as summarized in Table 9. 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 understanding of area models at the immediate and the delayed posttests.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Test</th>
<th>Fluency area with models</th>
<th>Fluency with number lines</th>
<th>Conceptual transfer</th>
<th>Procedural transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>fully interleaved &gt; blocked, moderately interleaved</td>
<td>post</td>
<td>-</td>
<td>ns</td>
<td>ns</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>delayed</td>
<td>-</td>
<td>ns</td>
<td>$p &lt; .05, d = .33$</td>
<td>-</td>
</tr>
<tr>
<td>increasingly interleaved &gt; blocked, moderately interleaved, fully interleaved</td>
<td>post</td>
<td>$p &lt; .10, d = .30$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>delayed</td>
<td>$p &lt; .10, d = .30$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>moderately interleaved &gt; blocked, fully interleaved</td>
<td>post</td>
<td>-</td>
<td>-</td>
<td>ns</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>delayed</td>
<td>-</td>
<td>-</td>
<td>ns</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 9. Results from post-hoc 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 post-hoc comparisons were computed.

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

**Effects of practice schedules on students' benefit from multiple graphical representations**

To investigate hypothesis 2 (that interleaved practice schedules enhance students’ benefit from multiple graphical representations compared to a single graphical representation), I applied the HLM described in formula (3) for the interleaved condition and the single-representation control conditions. I then computed planned contrasts that compared the 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 fluency with number lines at the immediate posttest, $t(445) = 2.09, p < .05, d = .09$, on fluency 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, $t(445) = 2.27, p < .05, d = .10$.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Effect</th>
<th>Fluency with area models</th>
<th>Fluency with number lines</th>
<th>Conceptual transfer</th>
<th>Procedural transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>blocked</td>
<td>post &gt; pre</td>
<td>ns</td>
<td>ns</td>
<td>$p &lt; .05, d = .42$</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>delayed &gt; pre</td>
<td>$p &lt; .05, d = .52$</td>
<td>$p &lt; .01, d = .39$</td>
<td>$p &lt; .05, d = .39$</td>
<td>ns</td>
</tr>
<tr>
<td>moderately interleaved</td>
<td>post &gt; pre</td>
<td>ns</td>
<td>ns</td>
<td>$p &lt; .05, d = .29$</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>delayed &gt; pre</td>
<td>$p &lt; .01, d = .50$</td>
<td>$p &lt; .05, d = .30$</td>
<td>$p &lt; .05, d = .45$</td>
<td>ns</td>
</tr>
<tr>
<td>fully interleaved</td>
<td>post &gt; pre</td>
<td>$p &lt; .05, d = .45$</td>
<td>$p &lt; .01, d = .51$</td>
<td>$p &lt; .01, d = .34$</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>delayed &gt; pre</td>
<td>$p &lt; .05, d = .38$</td>
<td>$p &lt; .01, d = .75$</td>
<td>$p &lt; .01, d = .60$</td>
<td>ns</td>
</tr>
<tr>
<td>increasingly interleaved</td>
<td>post &gt; pre</td>
<td>$p &lt; .05, d = .38$</td>
<td>$p &lt; .01, d = .43$</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>delayed &gt; pre</td>
<td>$p &lt; .05, d = .55$</td>
<td>$p &lt; .01, d = .46$</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>single-circle</td>
<td>post &gt; pre</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>delayed &gt; pre</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>single-rectangle</td>
<td>post &gt; pre</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>delayed &gt; pre</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>single-number-line</td>
<td>post &gt; pre</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>delayed &gt; pre</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
</tbody>
</table>

Table 10. Improvement of 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 fluency with number lines, $F(2, 867) = 20.09, p < .01$, partial $\eta^2 = .03$, for fluency 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 fluency with ar-
ea models $F(12, 862) = 2.06, p < .05, \text{partial } \eta^2 = .01$. These results show that students (regardless of condition) improved on fluency with number lines, fluency with area models, procedural and conceptual transfer, albeit with small effect sizes. On fluency with area models, students’ learning gains depended on the condition.

To further clarify these results, I computed post-hoc comparisons that contrasted students’ scores at the immediate posttest and the delayed posttest, compared to the pretest. Table 10 provides a summary of these post-hoc comparisons. Generally, I found significant learning gains at the delayed posttest for most of the multiple graphical representations conditions on understanding of area models, understanding of 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.

4.3.1.4. Summary

In summary, Experiment 3 shows 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. Furthermore, there was a marginally significant advantage for the increasingly interleaved condition over the blocked, moderately interleaved, and fully interleaved conditions on fluency with the number line at the immediate and the delayed posttests. Although a comparison of all multiple graphical representations conditions and the single graphical representation condition was not significant for the majority of dependent measures, the interleaved multiple graphical representations condition significantly outperformed the single graphical representation conditions on fluency with the number line, conceptual transfer, and marginally on procedural transfer.

Taken together, the findings on the learning outcomes provide some evidence that interleaved practice with multiple graphical representations (in addition to moderately interleaving task types) leads to better learning of fractions than blocked practice by promoting students’ representational fluency. Furthermore, only when provided in an interleaved sequence were multiple graphical representations more effective than a single graphical representation.
4.3.2 Think-aloud study to investigate learning processes

But how do we know that the advantage of interleaving graphical representations can be attributed to representational fluency, rather than to students’ connection making between multiple graphical representations across consecutive tutor problems – which would indicate that interleaved practice with multiple graphical representations supports students’ connectional understanding rather than their representational fluency?

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 11 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. Table 11 suggests 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 almost none of the conceptual connections. Explicit prompts seem to have yielded more of the conceptual connections, compared to the implicit prompts.

<table>
<thead>
<tr>
<th></th>
<th>Implicit prompts</th>
<th>Explicit prompts</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface</td>
<td>4.17</td>
<td>2.33</td>
<td>2.94</td>
</tr>
<tr>
<td>Conceptual</td>
<td>.58</td>
<td>1.63</td>
<td>1.28</td>
</tr>
<tr>
<td>Overall</td>
<td>2.38</td>
<td>1.98</td>
<td></td>
</tr>
</tbody>
</table>

Table 11. Number of surface connections and conceptual connections by implicit and explicit prompts averaged across students.

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.
These findings demonstrate 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 re-activate 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.

Based on the results from the think-aloud study, it seems likely that the advantage of interleaved practice on learning outcomes is not due to students’ conscious connection making between multiple graphical representations, but from repeated reactivation of knowledge about graphical representations, which, as argued, might help students in developing representational fluency.

### 4.3.3 Educational data mining to investigate learning processes

<table>
<thead>
<tr>
<th>Task type</th>
<th>Blocked</th>
<th>Moderate</th>
<th>Interleaved</th>
<th>Increased</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.85 (.08)</td>
<td>.86 (.06)</td>
<td>.86 (.06)</td>
<td>.87 (.05)</td>
</tr>
<tr>
<td>2</td>
<td>.88 (.07)</td>
<td>.89 (.07)</td>
<td>.89 (.05)</td>
<td>.89 (.05)</td>
</tr>
<tr>
<td>3</td>
<td>.91 (.05)</td>
<td>.88 (.07)</td>
<td>.88 (.06)</td>
<td>.87 (.07)</td>
</tr>
<tr>
<td>4</td>
<td>.87 (.06)</td>
<td>.84 (.07)</td>
<td>.81 (.07)</td>
<td>.84 (.06)</td>
</tr>
<tr>
<td>5</td>
<td>.83 (.10)</td>
<td>.86 (.06)</td>
<td>.83 (.08)</td>
<td>.83 (.07)</td>
</tr>
<tr>
<td>6</td>
<td>.88 (.10)</td>
<td>.90 (.06)</td>
<td>.89 (.07)</td>
<td>.89 (.07)</td>
</tr>
<tr>
<td>Overall</td>
<td>.87 (.08)</td>
<td>.87 (.07)</td>
<td>.86 (.07)</td>
<td>.87 (.06)</td>
</tr>
</tbody>
</table>

Table 12. Average number of correct first attempts by task type and condition (standard deviation in brackets). Higher numbers indicate higher performance during the acquisition phase.

Yet, if interleaved practice with graphical representations helps students in developing representational fluency, should not the tutor log data show that students in the interleaved condition outperform students in the blocked condition? Table 12 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 dif-
ferences in students’ performance ($t < 1$). Therefore, based on raw performance measures, I do not find evidence that interleaved practice with graphical representations promotes students’ development of fluency.

This lack of an advantage of the interleaved condition on correct attempts may not be surprising: a common finding in the literature on practice schedules 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. These findings indicate that raw measures of performance, such as average number of first attempts, may not be a suitable metric to detect differences between conditions during the acquisition phase. Yet, knowledge tracing, which tracks student knowledge over time, can be used to investigate learning differences between conditions during the acquisition phase (Corbett & Anderson, 1995; Pardos, Dailey, & Heffernan, 2010). Therefore, I investigated whether learning rate estimates, based on knowledge tracing, can detect the advantage of interleaved practice even during the acquisition phase (also see Rau & Pardos, 2012). Consequently, this analysis investigates whether knowledge tracing provides a suitable metric for detecting the effects of an intervention that is known not be accessible through simpler metrics such as performance 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.

I employed two models which 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. 18 provides an overview of the different models that we 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.

Fig. 18. Overview of the four different Bayesian Networks tested, with observed (o.) and hidden (h.) nodes.

In order to determine model fit by task type, I analyzed the log data by tasks type. For the evaluation of predictive performance, reported below, a 5-fold cross-validation at the student level was used. 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 is in the unlearned state, the probability that he/she will transition to the learned state between problems is:

\[
P(L_{n-1}) + ((1 - P(L_{n-1})) * P(T|C_s)),
\]

where \( P(L_{n-1}) \) is the probability of a student already knowing the skill, 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

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 results in greater 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.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition-Representation-Analysis Model</td>
<td>.3427</td>
<td>.6528</td>
</tr>
<tr>
<td>Standard-Knowledge-Tracing Model</td>
<td>.3445</td>
<td>.6181</td>
</tr>
<tr>
<td>Condition-Analysis Model</td>
<td>.3466</td>
<td>.5509</td>
</tr>
<tr>
<td>Prior-Per-Student Model</td>
<td>.3469</td>
<td>.5604</td>
</tr>
</tbody>
</table>

Table 13. Summary of the cross-validated prediction results of the four tested models using RMSE and AUC metrics.

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

Table 14. Learning rates by task type and condition from the Condition-Representation Analysis Model. Higher numbers indicate higher learning rates during the acquisition phase.
As shown in Table 13, 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 14 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 (namely 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 tasks 1, 2 and 3 ($p < .05$) and moderately significant for task 5 ($p < .10$). The interleaved condition achieved the highest overall learning rate which was twice that of any other condition. This advantage is remarkable, given that performance, as established by the average number of errors made during the acquisition phase (see Table 12), did not differ between conditions.

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 the best fit to the log data confirms the overall finding from the classroom experiment 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 confirms 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 finding from the posttest data that interleaved practice schedules of multiple graphical representations of fractions lead to better learning than blocked practice schedules. This finding is interesting, especially in the light of the lack of statistically significant differences in students’ performance on the Fractions Tutor during the acquisition phase. As shown in Table 12, 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 temporal variation 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 the 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 confirm that interleaving graphical representations leads to better learning than blocking graphical representations and demonstrate that the advantage of interleaved practice is due to enhancing representational fluency-building processes as students learn to use individual graphical representations more efficiently to solve fractions problems.

The finding that the advantage of interleaved practice can be identified during the acquisition phase is particularly interesting against the background of prior research on practice schedules. A vast amount of prior research documents that interleaved practice decreases 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). The Bayesian knowledge tracing analysis demonstrates that face-value methods, such as performance measures during the acquisition phase as typically used in prior research, do not provide sufficient information to evaluate an educational intervention. This finding is in line with the notion that interventions that improve performance do not necessarily improve learning, but that interventions which decrease performance by making the learning task more difficult, can improve learning (Schmidt & Bjork, 1992) – a well-researched phenomenon known as desirable difficulties (McDaniel & Butler, 2010; Paas & Van Merrienboër, 1994; Rohrer & Pashler, 2010) and as the assistance dilemma (Kapur & Rummel, 2009; Koedinger & Aleven, 2007; Koedinger, Pavlik, McLaren, & Aleven, 2008). The findings from the Bayesian knowledge tracing analysis contradict the common notion that the advantages of interventions (such as interleaved practice) which decrease performance is not apparent during the acquisition phase. Learning rate estimates based on Bayesian knowledge tracing provide an assessment of learning during the acquisition phase that looks beyond performance measures.

Since many domains use multiple graphical representations to augment instructional materials, I believe that these findings have the potential to generalize across a wide range of learning materials. Bayesian knowledge tracing analyses can help us make sense of the complex educational data that we obtain from the rich settings in which education takes place, and hence, help us understand complex learning processes.

The finding from Experiment 3, that one should interleave practice with multiple graphical representations, does not contradict the finding from Experiment 2, that one should interleave task types rather than multiple graphical representations. Experiment 3 built on the outcomes of
Experiment 2 in that a moderately interleaved sequence of task types was used consistently across conditions. Experiment 3 therefore answered the question whether in addition to interleaving task types, multiple graphical representations should also be interleaved. Furthermore, interleaved sequences of multiple graphical representations and task types serve different purposes: interleaving task types serves the development of *representational understanding*, whereas interleaving multiple graphical representations serves the development of *representational fluency*. 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: 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.
4.4 Experiment 4: Supporting connectional sense-making and fluency-building processes

Students need not only to develop understanding and fluency with individual representations, but also of the connections between them. The goal of Experiment 4 was to investigate the complementary role of two learning processes involved in connection making between multiple graphical representations: connectional sense-making processes and connectional fluency-building processes. Connection making between multiple representations is considered key to students’ benefit from multiple representations (Ainsworth, 2006; Cook et al., 2007; Even, 1998; Gutwill et al., 1999; Özugün-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). Yet, 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). As described in section 2.2, I expect that both connectional sense-making processes and connectional fluency-building processes play a role in connection making between 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 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.

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

Another question that the literature on connection making support leaves open is the role of automated support provided by the system for connectional sense-making processes. 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: worked examples have not yet been investigated as a means of support for connectional sense-making processes involving 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 automatically by the educational technology. A further goal of Experiment 4 is therefore to investigate how much automated support for connectional sense-making processes students should receive 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:_ Students’ acquisition of robust knowledge of fractions requires support for both connectional sense-making and fluency-building processes.

_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 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 15. Specifically, 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.

<table>
<thead>
<tr>
<th>Fluency-building support</th>
<th>Sense-making support</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>None</td>
<td>Control</td>
</tr>
<tr>
<td>Auto-linked representations condition</td>
<td>Auto-linked condition</td>
<td>Fluency-building condition</td>
</tr>
<tr>
<td>Worked example condition</td>
<td>Number-line condition</td>
<td></td>
</tr>
<tr>
<td>Fluency-building condition</td>
<td>Auto-linking and fluency-building condition</td>
<td>Worked-example and fluency-building condition</td>
</tr>
</tbody>
</table>

Table 15. Overview of experimental conditions in Experiment 4.

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 which 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, \(\chi^2 (6, N = 169) = 4.34, p > .10\).
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 at all times 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, we gave students an equivalent delayed posttest.

![Diagram of the procedure for conditions with connection-making support](image1)

**Fig. 19.** Procedure for conditions with connection-making support.

![Diagram of the procedure for conditions with sense-making and fluency-building support](image2)

**Fig. 20.** Procedure for conditions with sense-making and fluency-building support.
Fig. 19 illustrates this procedure for students in any of the conditions with connection-making support (i.e., auto-linked condition, worked-example condition, fluency-building condition, auto-linking and fluency-building condition, worked-example and fluency-building condition). For each of the topics covered by the Fractions Tutor, students first completed four single-representation tutor problems, followed by four connection-making problems, assigned according to the student’s condition. 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. 20. 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.

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. 21 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. 21) which asked students to identify correspondences of the two given representations.
4.4.2.5. Test instruments

Students took three tests: a pretest, an immediate posttest, and a delayed posttest. 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. Examples of the test items can be found in Appendix 6. 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. 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. I validated the resulting factor structure using the data from the immediate and the delayed posttests.

### 4.4.3 Results

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

### Table 16. Proportion correct: means (and standard deviation) for conceptual and procedural knowledge at pretest, immediate posttest, delayed posttest. Min. score is 0, max. score is 1.

<table>
<thead>
<tr>
<th>measure</th>
<th>pretest</th>
<th>immediate posttest</th>
<th>delayed posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>multiple graphical representations</td>
<td>0.33 (0.20)</td>
<td>0.45 (0.23)</td>
<td>0.48 (0.26)</td>
</tr>
<tr>
<td>auto-linking</td>
<td>0.38 (0.20)</td>
<td>0.49 (0.23)</td>
<td>0.51 (0.26)</td>
</tr>
<tr>
<td>worked examples</td>
<td>0.36 (0.22)</td>
<td>0.43 (0.20)</td>
<td>0.49 (0.26)</td>
</tr>
<tr>
<td>fluency-building</td>
<td>0.31 (0.21)</td>
<td>0.37 (0.22)</td>
<td>0.44 (0.24)</td>
</tr>
<tr>
<td>auto-linking and fluency-building</td>
<td>0.36 (0.20)</td>
<td>0.43 (0.24)</td>
<td>0.49 (0.25)</td>
</tr>
<tr>
<td>worked examples and fluency-building</td>
<td>0.39 (0.21)</td>
<td>0.52 (0.24)</td>
<td>0.58 (0.26)</td>
</tr>
<tr>
<td>number-line only</td>
<td>0.37 (0.20)</td>
<td>0.43 (0.25)</td>
<td>0.48 (0.20)</td>
</tr>
</tbody>
</table>

### Table 17. Overview of results from Experiment 4.
As mentioned, I analyzed the data of \( N = 428 \) students. Table 16 shows the means and standard deviations for the conceptual and procedural knowledge scales by test time and condition. Table 17 gives an overview of the results.

### 4.4.3.1. Effects of connection-making support on learning outcomes

To take into account nested sources of variability due to students being in different classes and in different schools, a hierarchical linear model (Raudenbush & Bryk, 2002) with four nested levels was used to analyze the data. I modeled performance for each of the three tests for each student (level 1), differences between students (level 2), differences between classes (level 3), and between schools (level 4). More specifically, the following HLM was fitted to the data:

\[
Y_{ijkl} = (((\mu_{0000} + W_{000l}) + V_{00kl}) + \beta_3 * f_{ijkl} + \beta_2 * s_{ijkl} + U_{ijkl}) + \beta_1 * t_{ijkl} + R_{ijkl}
\]

with

(1) \( Y_{ijkl} = \beta_{ijkl} + \beta_1 * t_{ijkl} + R_{ijkl} \)

(2) \( \beta_{ijkl} = \delta_{00kl} + p_{ijkl} + \beta_2 * s_{ijkl} + \beta_3 * f_{ijkl} + \beta_4 * p_{ijkl} * s_{ijkl} + \beta_5 * p_{ijkl} * f_{ijkl} \)

(3) \( \delta_{00kl} = \gamma_{000l} + V_{00kl} \)

(4) \( \gamma_{000l} = \mu_{0000} + W_{000l} \)

where the dependent variable \( Y_{ijkl} \) is student \( i \)'s score on the dependent measures at test \( t_i \) (i.e., immediate or delayed posttest), \( \beta_{ijkl} \) is student \( i \)'s score across tests, \( \beta_1 \) is the effect of test time, \( \delta_{ijkl} \) is the average performance of class \( k \), \( p_{ijkl} \) is student \( j \)'s performance on the pretest, \( \beta_3 \) being the effect of sense-making support, \( \beta_4 \) being the effect of fluency-building support, and \( \gamma_{000l} \) is the average performance of school \( l \), and \( \mu_{0000} \) is the overall average. I included the interaction of students’ pretest scores with sense-making support and with fluency-building support (\( \beta_4 * p_{ijkl} * s_{ijkl} \), and \( \beta_5 * p_{ijkl} * f_{ijkl} \)) to investigate aptitude-treatment interaction effects (i.e., whether students’ benefits from one condition over the other depends on their prior knowledge).

Since the HLM described in (5) 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 included pretest score in the dependent variable, yielding:

\[
Y_{ijkl} = (((\mu_{0000} + W_{000l}) + V_{00kl}) + \beta_3 * f_{ijkl} + \beta_2 * s_{ijkl} + U_{ijkl}) + \beta_6 * t_{ijkl} * s_{ijkl} + \beta_5 * t_{ijkl} * f_{ijkl} + \beta_1 * t_{ijkl} + R_{ijkl}
\]

with

(1) \( Y_{ijkl} = \beta_{ijkl} + \beta_6 * t_{ijkl} + \beta_5 * t_{ijkl} * s_{ijkl} + \beta_5 * t_{ijkl} * f_{ijkl} + R_{ijkl} \)

(2) \( \beta_{ijkl} = \delta_{00kl} + p_{ijkl} + \beta_2 * s_{ijkl} + \beta_3 * f_{ijkl} + \beta_4 * p_{ijkl} * s_{ijkl} + \beta_5 * p_{ijkl} * f_{ijkl} \)}
(level 3) $\delta_{00kl} = \gamma_{000l} + V_{00kl}$

(level 4) $\gamma_{000l} = \mu_{0000} + W_{000l}$

where the dependent variable $Y_{ijkl}$ is student $j$’s score on the dependent measures at test $i$ (i.e., pretest, immediate posttest, or delayed posttest), $\beta_5$ being the effect of the interaction of sense-making support with test time, $\beta_6$ being the effect of the interaction of fluency-building support with test time.

The reported $p$-values were adjusted for multiple comparisons using the Bonferroni correction. I report partial $\eta^2$ for effect sizes on main effects and interactions between factors, and Cohen’s $d$ for effect sizes of pairwise comparisons. An effect size partial $\eta^2$ 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.

**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 formula (4). 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 confirm the hypothesis that both types of support are needed for conceptual knowledge: 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 *post-hoc* 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 results show that sense-making processes and fluency-building processes in connection making both need to be supported in order for students’ conceptual knowledge of fractions 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 formula (4): a test of the effect of sense-making support for each level of the fluency support factor. Effect slices for the effect of 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 ($F < 1$). Post-hoc 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 both connectional sense-making and fluency-building support are needed in order to promote students’ conceptual learning of fractions. 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 both on connectional sense-making and fluency-building processes.
### 4.4.3.2. Causal path modeling

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Description</th>
<th># in worked-example condition</th>
<th># in fluency-building condition</th>
<th># in worked-example and fluency-building condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>additionMixedError</td>
<td>Finding representations that show the addend of a given sum equation depicted by representations</td>
<td>n/a</td>
<td>207</td>
<td>176</td>
</tr>
<tr>
<td>compareMixed-Error</td>
<td>Finding representations that show a fraction smaller or larger than the given one</td>
<td>n/a</td>
<td>436</td>
<td>307</td>
</tr>
<tr>
<td>comparisonError</td>
<td>Comparing two fractions</td>
<td>92</td>
<td>n/a</td>
<td>82</td>
</tr>
<tr>
<td>denomError</td>
<td>Entering the denominator of a fraction</td>
<td>972</td>
<td>n/a</td>
<td>837</td>
</tr>
<tr>
<td>DiffMixedError</td>
<td>Finding representations that show the difference of two fractions</td>
<td>n/a</td>
<td>282</td>
<td>238</td>
</tr>
<tr>
<td>equivalence-CompareError</td>
<td>Finding equivalent fraction representations</td>
<td>n/a</td>
<td>2899</td>
<td>2157</td>
</tr>
<tr>
<td>equivalenceError</td>
<td>Finding representations of improper fractions</td>
<td>n/a</td>
<td>1380</td>
<td>1608</td>
</tr>
<tr>
<td>multiplyError</td>
<td>Entering a number by which to multiply numerator or denominator to expand a given fraction</td>
<td>30</td>
<td>n/a</td>
<td>29</td>
</tr>
<tr>
<td>nameCircleMixedError</td>
<td>Finding circle representations that show the same fraction as a number line or a rectangle</td>
<td>n/a</td>
<td>355</td>
<td>126</td>
</tr>
<tr>
<td>nameNLMixedError</td>
<td>Finding number line representations that show the same fraction as a circle or a rectangle</td>
<td>n/a</td>
<td>949</td>
<td>599</td>
</tr>
<tr>
<td>nameRectMixedError</td>
<td>Finding rectangle representations that show the same fraction as a number line or a circle</td>
<td>n/a</td>
<td>385</td>
<td>133</td>
</tr>
<tr>
<td>nlPartitionError</td>
<td>Partitioning the number line to show an equivalent fraction</td>
<td>1913</td>
<td>n/a</td>
<td>2115</td>
</tr>
<tr>
<td>numberSections-UnitError</td>
<td>Finding the denominator of a fraction by indicating how many sections the unit was divided into</td>
<td>41</td>
<td>n/a</td>
<td>44</td>
</tr>
<tr>
<td>numError</td>
<td>Entering the numerator of a fraction</td>
<td>1559</td>
<td>n/a</td>
<td>1390</td>
</tr>
</tbody>
</table>

6 Significant difference between conditions, based on chi-square test
7 Significant predictor of performance on conceptual posttest, after controlling for pretest performance
### Table 18. Error types and number of error types per condition. Error types in italics were selected for further analysis.

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Description</th>
<th>Count</th>
<th>NA</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>place1Error^a/c</td>
<td>Locating 1 on the number line given a dot on the number line and the fraction it shows</td>
<td>150</td>
<td>n/a</td>
<td>222</td>
</tr>
<tr>
<td>placeDotError^a</td>
<td>Placing a dot on the number line to show a fraction</td>
<td>198</td>
<td>n/a</td>
<td>253</td>
</tr>
<tr>
<td>sectionsBetween-0-1</td>
<td>Indicating that the denominator in a number line is shown by the sections between 0 and 1</td>
<td>61</td>
<td>n/a</td>
<td>44</td>
</tr>
<tr>
<td>SE-Error^a/c</td>
<td>Self-explanation error, response to reflection questions</td>
<td>1320</td>
<td>n/a</td>
<td>1629</td>
</tr>
<tr>
<td>subtractionMixed-Error^d</td>
<td>Finding representations that show the subtrahend of a given difference equation depicted by representations</td>
<td>n/a</td>
<td>214</td>
<td>240</td>
</tr>
<tr>
<td>sumMixedError^d</td>
<td>Finding representations that show the sum of two fractions</td>
<td>n/a</td>
<td>256</td>
<td>205</td>
</tr>
<tr>
<td>unitError</td>
<td>Selecting the unit for a fraction given the symbolic fraction and a graphical representation</td>
<td>123</td>
<td>n/a</td>
<td>115</td>
</tr>
<tr>
<td>unitMixedError</td>
<td>Finding the unit of a given fraction</td>
<td>n/a</td>
<td>1050</td>
<td>1138</td>
</tr>
</tbody>
</table>

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 support condition, fluency-building support condition, and combined worked-example and fluency-building support conditions made on the worked example problems and on the fluency-building problems. I compared the frequency of error types on those connection-making problems that were the same for two given conditions: errors on worked examples problems in the worked-example condition and the worked-example support conditions, and errors on fluency-building problems in the fluency-building condition and the worked-example and fluency-building conditions.
building condition. Table 18 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. 22 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.

Fig. 23 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. 22. Structural equation model for the worked-example condition and the worked-example and fluency-building condition.](image)

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. 23. Structural equation model for the fluency-building condition and the worked-example and fluency-building condition.](image)

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 delayed posttest.
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 demonstrates that both connectional sense-making processes and fluency-building processes are necessary for students to benefit from multiple graphical representations: only students who received both types of instructional support for connection making significantly outperformed a single-representation control condition. Thereby, Experiment 4 extends previous research on connection making that has (1) only focused on either connectional sense-making processes or connectional fluency-building processes but not on the combination of 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.

Specifically, the findings from Experiment 4 show that connectional sense-making support between multiple graphical representations is effective, as long as it is combined with connectional fluency-building support. Furthermore, students need to become active in engaging in connectional sense-making processes, rather than observing connections provided by auto-linked representations. In addition, Experiment 4 demonstrates that support for connectional fluency-building processes is effective as long as combined with sense-making support for connection making.

Taken together, Experiment 4 shows that connectional sense-making processes and fluency-building processes interact. But, it remains unclear how these learning processes interact. The causal path analysis on the tutor log data obtained from Experiment 4 leads to the hypothesis that
connectional understanding provides the foundation for students’ benefit from support for con-
nectional fluency-building processes. Experiment 5 was designed to investigate this hypothesis.
Experiment 4 shows that both connectional sense-making processes and connectional fluency-building processes need to be supported for students to benefit from multiple graphical representations. While it is evident that connectional sense-making processes and fluency-building processes interact, it remains unclear how they interact. Does connectional understanding enhance fluency-building processes, or does connectional fluency enhance connectional sense-making processes? 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 learning. Based on an structural equation model of students’ error types, one might hypothesize that connectional sense-making support helps students 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 processes and fluency-building processes interact is also of practical relevance. If connectional understanding (acquired through 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 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 processes and fluency-building processes interact, my research is a step towards closing the gap between studies that have exclusively 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). In this section, I describe theoretical perspectives pertaining to 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 engage in to master a domain. And 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 – they have recently put more emphasis on fluency as well. 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 fluency in making connections between graphical representations. Yet, the NMAP (2008) describes fluency in relating different fractions representations as one important foundation for later algebra learning. Unfortunately, neither of these literatures makes explicit claims about dependences between sense-making processes and fluency-building processes, thus obviating the need for an empirical investigation of this question. In the following, I 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 which can be assessed 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 which might be missing for the completion of a more complex task, such as tasks that involve connectional sense-making processes. Is this the case, providing connectional sense-making support before connectional fluency-building support may enhance chances 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. The ability to efficiently work with fractions representations is expected later – not before the end of grade 8.

Instructional design principles based on Cognitive Load theory provide a theoretical rationale for the understanding-first hypothesis. Connectional fluency-building support involves perceptually rich learning tasks while providing minimal guidance for students to solve them. Students are expected to learn (through discovery) to extract structurally relevant information from experience across a variety of representations. Kirschner, Sweller, and Clark (2006) vigorously argue that minimally-guided practice with information-rich problems may increase cognitive load during problem solving and may hamper learning. In vivid terms, they describe that “minimally-guided learning does not enhance student achievement any more than throwing a non-swimmer out of a boat in the middle of the lake supports learning to swim” (p. 4).

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 which 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 on learning outcomes

First, 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 necessary to benefit from connectional fluency-building support, for instance by helping them 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, 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, 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 which 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 on problem-solving behaviors**

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 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 on connection making and conceptual processing

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. By contrast, the fluency-first condition does not predict differences between conditions on connection making while students 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 working 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 working on fluency-building problems.
Furthermore, the fluency-first condition 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 working on connectional sense-making problems. However, the understanding-first condition does not predict differences between conditions on connection making while students 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 working 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 working on 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 make engage in more conceptual reasoning than the fluency-first condition. 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 than the fluency-first condition.

**Fluency-first concepts H2:** The fluency-first condition will engage in more conceptual reasoning about fractions 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.

**Predictions on visual attention**

Both the understanding-first and the fluency-first hypotheses also make predictions about students' visual attention during the learning process. According to the understanding-first hypothe-
sis, students in the understanding-first condition should demonstrate more eye-gaze behaviors that are indicative of integrative processing and conceptual reasoning while working on connectional fluency-building support than students in the fluency-first condition, because their connectional understanding prepares them to take better advantage of connectional fluency-building support. The fluency-first condition, on the other hand, does not predict differences between conditions in eye-gaze behaviors while students work on connectional fluency-building problems, because students' benefit from connectional fluency-building support is considered as independent from connectional understanding. Thus, I contrast the following two hypotheses:

**Understanding-first eye-gaze-fluency H1:** The understanding-first condition will exhibit more eye-gaze behaviors that are considered to indicate integrative processing and conceptual reasoning while working on connectional fluency-building support.

**Fluency-first eye-gaze-fluency H0:** The fluency-first condition and the understanding-first condition will exhibit the same amount of eye-gaze behaviors that are considered to indicate integrative processing and conceptual reasoning while working on connectional fluency-building support.

By contrast, the fluency-first condition predicts that connectional fluency equips students with the prerequisite head-room to take full advantage of connectional sense-making support. Therefore, students in the fluency-first condition should demonstrate more eye-gaze behaviors indicative of integrative processing and conceptual reasoning than students in the understanding-first condition while working on connectional sense-making support. By contrast, the understanding-first condition predicts that students should be able to engage in integrative processing and conceptual reasoning independently of previously acquired connectional fluency and therefore does not predict differences between conditions on such eye-gaze behaviors while students work on connectional sense-making support. Thus, I compare the following two hypotheses:

**Fluency-first eye-gaze-sense H1:** The fluency-first condition will exhibit more eye-gaze behaviors that are considered to indicate integrative processing and conceptual reasoning while working on connectional sense-making support.

**Understanding-first eye-gaze-sense H0:** The understanding-first condition and the fluency-first condition will exhibit the same amount of eye-gaze behaviors that are considered to indicate
integrative processing and conceptual reasoning while working on connectional sense-making support.

Predictions about 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

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. 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.
4.5.2.2. Participants
Seventy-four students from grades 3-5 participated in the experiment. Sessions were conducted individually in the lab. Students were randomly assigned to the understanding-first condition or to the fluency-first condition.

4.5.2.3. Procedure
Table 19 details the sequence of assessment problems and tutor problems for each experimental condition. Students first completed a pretest. They then worked on 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.

<table>
<thead>
<tr>
<th>Activity Type</th>
<th>Understanding-first condition</th>
<th>Fluency-first condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>Pretest: near / far transfer</td>
<td>Pretest: near / far transfer</td>
</tr>
<tr>
<td>Tutor: equivalence</td>
<td>Sense-making support: 4 tutor problems</td>
<td>Fluency-building support: 4 tutor problems</td>
</tr>
<tr>
<td>Quiz 1: equivalence</td>
<td>Reproduction-understanding, reproduction-fluency</td>
<td>Reproduction-understanding, reproduction-fluency</td>
</tr>
<tr>
<td>Tutor: equivalence</td>
<td>Fluency-building support: 4 tutor problems</td>
<td>Sense-making support: 4 tutor problems</td>
</tr>
<tr>
<td>Quiz 2: equivalence</td>
<td>Reproduction-understanding, reproduction-fluency</td>
<td>Reproduction-understanding, reproduction-fluency</td>
</tr>
<tr>
<td>Tutor: comparison</td>
<td>Sense-making support: 4 tutor problems</td>
<td>Fluency-building support: 4 tutor problems</td>
</tr>
<tr>
<td>Quiz 1: comparison</td>
<td>Reproduction-understanding, reproduction-fluency</td>
<td>Reproduction-understanding, reproduction-fluency</td>
</tr>
<tr>
<td>Tutor: comparison</td>
<td>Fluency-building support: 4 tutor problems</td>
<td>Sense-making support: 4 tutor problems</td>
</tr>
<tr>
<td>Quiz 2: comparison</td>
<td>Reproduction-understanding, reproduction-fluency</td>
<td>Reproduction-understanding, reproduction-fluency</td>
</tr>
<tr>
<td>Test</td>
<td>Posttest: near / far transfer</td>
<td>Posttest: near / far transfer</td>
</tr>
</tbody>
</table>

Table 19. Sequence of activities by experimental condition.

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 way as students in phase 1 (after following a calibration procedure that took about 1-2 minutes to complete).

4.5.2.4. Fractions Tutor versions

Students worked with a subset of the tutor problems used in Experiment 4 (also 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 problem and are then presented with an equivalent problem in which they have to use the number line to complete the problem themselves. At the end of the problem, students were prompted to relate the two graphical representations to one another. Fig. 10 (see section 3.4.2) shows an example of a connectional sense-making support problem for equivalent fractions, Fig. 11 (see section 3.4.2) shows an example of a connectional sense-making support problem for fraction comparison.

Connectional fluency-building support problems are based on 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. 12, see section 3.4.3), or order them from smallest to largest, using drag-and-drop (see Fig. 12, 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, eye-tracking measures, and conducted retrospective interviews.

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. Specifically, connection-understanding items assessed students' conceptual understanding of connections between graphical representations with regard to equivalent fractions and fraction
comparison. Connection-fluency items assessed students' fluency in making connections with regard to equivalent fractions and fraction comparison. 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):

\[
\text{efficiency} = \frac{Z(\text{proportion correct}) - Z(\text{time on quiz items})}{\sqrt{2}}
\]  
(7)

Higher efficiency scores indicate higher efficiency at solving quiz items correctly.

I assessed students' transfer of fractions knowledge based on equivalent pretests and post-tests. The transfer test included test items on equivalence and comparison without graphical representations. Example test items are provided in Appendix 7. I computed accuracy and efficiency scores for the transfer 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 20. Fig. 24 and 25 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 which are 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., 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 opportunities. 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 20), (2) equivalence knowledge components in sense-making problems (i.e., knowledge components 4-10 in Table 20), (3) comparison knowledge components in fluency-building problems (i.e., knowledge components 11-13 in Table 20), (4) comparison knowledge components in sense-making problems (i.e., knowledge components 14-21 in Table 20). 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 20.

![Example knowledge components for equivalent fraction sense-making problems.](image)
Fig. 25. Example knowledge components for fraction comparison sense-making problems.

<table>
<thead>
<tr>
<th>Knowledge component</th>
<th>Knowledge component type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. equivDrag_circle</td>
<td>Equivalence fluency-building</td>
<td>Dragging-and-dropping equivalent circle (fluency-building problem)</td>
</tr>
<tr>
<td>2. equivDrag_rect</td>
<td>Equivalence fluency-building</td>
<td>Dragging-and-dropping equivalent rectangle (fluency-building problem)</td>
</tr>
<tr>
<td>3. equivDrag_NL</td>
<td>Equivalence fluency-building</td>
<td>Dragging-and-dropping equivalent number line (fluency-building problem)</td>
</tr>
<tr>
<td>4. equivFractEquivalent</td>
<td>Equivalence sense-making</td>
<td>Judging whether two fraction representations are equivalent (sense-making problem)</td>
</tr>
<tr>
<td>5. equivMultiply</td>
<td>Equivalence sense-making</td>
<td>Multiplying the numerator or denominator of a given symbolic fraction to find an equivalent fraction (sense-making problem)</td>
</tr>
<tr>
<td>6. equivNameDenomFract</td>
<td>Equivalence sense-making</td>
<td>Entering the denominator of an equivalent fraction (sense-making problem)</td>
</tr>
<tr>
<td>7. equivNameNumFract</td>
<td>Equivalence sense-making</td>
<td>Entering the numerator of an equivalent fraction (sense-making problem)</td>
</tr>
<tr>
<td>8. relationEquivDiffNumbers</td>
<td>Equivalence sense-making</td>
<td>Reasoning about equivalent fractions showing the same amount with different numbers (sense-making problem)</td>
</tr>
<tr>
<td>9. relationEquivMultiplySameNumber</td>
<td>Equivalence sense-making</td>
<td>Reasoning about multiplying numerator and denominator by the same number to find equivalent fractions (sense-making problem)</td>
</tr>
<tr>
<td>10. relationEquivSameAmount</td>
<td>Equivalence sense-making</td>
<td>Reasoning about equivalent fractions showing the same amount (sense-making problem)</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Knowledge Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. compDrag_circle</td>
<td>Dragging-and-dropping circle that is smaller or larger than another (fluency-building problem)</td>
</tr>
<tr>
<td>12. compDrag_rect</td>
<td>Dragging-and-dropping rectangle that is smaller or larger than another (fluency-building problem)</td>
</tr>
<tr>
<td>13. compDrag_NL</td>
<td>Dragging-and-dropping number line that is smaller or larger than another (fluency-building problem)</td>
</tr>
<tr>
<td>14. compFract</td>
<td>Indicating which fraction representation is larger or smaller than another (sense-making problem)</td>
</tr>
<tr>
<td>15. compNumSect</td>
<td>Comparing the number of sections two fraction representations are partitioned into (sense-making problem)</td>
</tr>
<tr>
<td>16. compSectSize</td>
<td>Comparing the relative size of sections in two fraction representations (sense-making problem)</td>
</tr>
<tr>
<td>17. numSectZeroDot</td>
<td>Entering the fraction shown on the number line (sense-making problem)</td>
</tr>
<tr>
<td>18. relationCompLargerSizeLargerFract</td>
<td>Reasoning about the fraction representation with the larger sections showing the larger fraction, given that the numerators are the same (sense-making problem)</td>
</tr>
<tr>
<td>19. relationCompNumSectSizeSect</td>
<td>Reasoning about the number of sections a fraction representation is partitioned into being inversely related to the size of the sections (sense-making problem)</td>
</tr>
<tr>
<td>20. relationCompSameNum</td>
<td>Reasoning about the number of sections of two fraction representations being the same if the numerators are the same (sense-making problem)</td>
</tr>
<tr>
<td>21. relationCompTotalSectNumber</td>
<td>Reasoning about the number of total sections two fraction representations are partitioned into (sense-making problem)</td>
</tr>
</tbody>
</table>

**Eye-tracking measures**

As mentioned, I investigate eye-gaze behaviors that are considered to indicate integrative processing and conceptual reasoning. Educational psychology research has used eye-gaze measures to evaluate whether students pay attention to specific aspects of the learning content (Anderson & Gluck, 2001), to disambiguate strategy use (Anderson & Gluck, 2001), and to investigate whether students engage in sense-making activities such as self-explaining (Conati, Merten, Muldner, & Ternes, 2005). Several authors have argued that eye tracking is a useful methodology to gain insight into cognitive processes underlying the effect of instructional materials (Byerly, 2007; Hyönä, 2010; Jenkinson, 2009; Van Gog & Scheiter, 2010). Recently, eye-
tracking methods have also been increasingly used in multimedia learning research (Hyönä, 2010; Jenkinson, 2009; Johnson & Mayer, 2012; Mason et al., 2013; Mayer, 2010; Ozcelik et al., 2009; Schmidt-Weigand, Kohnert, & Glowalla, 2010a, 2010b; Schwonke & Renkl, 2010; Slykhuis, Wiebe, & Annetta, 2005; Thomas & Lleras, 2007; Van Gog & Scheiter, 2010), thus providing extensive guidance as to which eye-gaze behaviors are suitable to investigate learning processes. Although we have to be careful not to blindly trust eye-gaze behaviors as a direct mapping of cognitive processing (Irwin, 2004), the use of eye-gaze behaviors measures has enabled researchers to test predictions of theoretical accounts about cognitive processes fostered on how students visually integrate graphical representations and text (Hyönä, 2010; Mayer, 2010; Schmidt-Weigand et al., 2010a, 2010b), to investigate the effect of prior knowledge on perceptual processing (Carmichael, Larson, Gire, Loschky, & Rebello, 2010; Heo, Canham, & Fabrikant, 2010), and the effect of instructional interventions on students' visual attention (de Koning, Tabbers, Rikers, & Paas, 2010; Jarodzka, van Gog, Dorr, Scheiter, & Gerjets, 2013; Mason et al., 2013; Ozcelik et al., 2009), to mention only a few examples.

Eye-gaze behaviors that have typically been used in this prior research to indicate perceptual integration and cognitive processing are the frequency of switching between areas of interest (AOIs; Holsanova & Holmberg, 2009; Johnson & Mayer, 2012) and the duration of fixation after the first inspection of an AOI (Hyönä & Lorch, 2004; Hyönä, Lorch, & Rinck, 2003; Mason et al., 2013). The frequency of switching between AOIs is considered to indicate the mental integration of information presented in different AOIs. The first inspection of an AOI is considered to indicate an initial processing of the material which occurs (to a certain extent) automatically. The duration of fixations after the first inspection, that is, when a student reinspects an AOI after the first initial view, is considered to reflect intentional processing to integrate the information presented in the AOI with other information.

For the purpose of the current study, I created AOIs for each graphical and symbolic representation presented in each of the Fractions Tutor problems. Fig. 26 shows an example of the AOIs for a sense-making problem, Fig. 27 shows an example of the AOIs defined for a fluency-building problem. I computed the frequency of switching between these graphical representations as the number of consecutive fixations on different AOIs. I computed the duration of se-
cond-inspection fixations on each AOI as the sum of fixations that occurred after the first, initial fixation on AOIs for area models (i.e., circle or rectangle), and number lines.

**Fig. 26.** Connectional sense-making problems with areas of interest (AOIs) specified for area models, number lines, and symbolic fractions.

**Fig. 27.** Connectional fluency-building problem with areas of interest (AOIs) specified for area models and number lines.

**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. Specif-
ically, 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 20) 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. 24), 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 tracking), 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 way 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. 24], 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). For example, the interviewer might ask, after replaying the eye-gaze recording for step B2 in Fig. 24, “how did you solve this step?”

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 21 details the criteria for each of these codes. Fig. 28 shows a 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.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Representation connections</td>
<td>Only for sense-making problems, 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)</td>
</tr>
<tr>
<td>1.1. Representation-surface</td>
<td>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: That one (points at circle) has the same color as the rectangle.</td>
</tr>
<tr>
<td>1.2. Representation-concept-incorrect</td>
<td>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: That one (points at circle) has three shaded and that rectangle also has three shaded, so the circle is bigger.</td>
</tr>
<tr>
<td>1.3. Representation-concept-correct</td>
<td>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: That one (points at circle) has three shaded and that rectangle also has three shaded, but the circle sections are larger, so the circle is bigger.</td>
</tr>
<tr>
<td>2. Representation-fluency</td>
<td>Only for fluency-building problems, 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: The circle kind of looked like it’s the same size as the rectangle.</td>
</tr>
<tr>
<td>3. Concept-correct</td>
<td>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: When there are three out of five and three out of seven, the three out of five is larger because the parts are bigger.</td>
</tr>
</tbody>
</table>

Table 21. 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. 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 $\eta^2$ of .01 corresponding to a small effect, .06 to a medium effect, and .14 to a large effect (Cohen, 1988). Table 22 gives an overview of the results on the learning outcome measures. Table 23 provides an overview of the results on the process-level measures.
<table>
<thead>
<tr>
<th>measure</th>
<th>test time</th>
<th>effect</th>
<th>significant</th>
<th>F/t-value</th>
<th>adj. p-value</th>
<th>effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiz: connection-fluency accuracy</td>
<td></td>
<td>condition</td>
<td>marginally</td>
<td>(F(1,65) = 3.34)</td>
<td>(p &lt; .10)</td>
<td>(\eta^2 = .05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>quiz time</td>
<td>no</td>
<td>(F &lt; 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>condition * quiz time</td>
<td>no</td>
<td>(F(1,65) = 1.42)</td>
<td>(p &gt; .10)</td>
<td></td>
</tr>
<tr>
<td>quiz 1</td>
<td></td>
<td>understanding-first &gt; fluency-first</td>
<td>no</td>
<td>(F &lt; 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quiz 2</td>
<td></td>
<td>understanding-first &gt; fluency-first</td>
<td>yes</td>
<td>(F(1,65) = 4.52)</td>
<td>(p &lt; .05)</td>
<td>(\eta^2 = .07)</td>
</tr>
<tr>
<td>Quiz: connection-fluency efficiency</td>
<td></td>
<td>condition</td>
<td>yes</td>
<td>(F(1,65) = 12.14)</td>
<td>(p &lt; .01)</td>
<td>(\eta^2 = .16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>quiz time</td>
<td>no</td>
<td>(F(1,65) = 1.17)</td>
<td>(p &gt; .10)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>condition * quiz time</td>
<td>yes</td>
<td>(F(1,65) = 6.55)</td>
<td>(p &lt; .05)</td>
<td>(\eta^2 = .09)</td>
</tr>
<tr>
<td>quiz 1</td>
<td></td>
<td>fluency-first &gt; understanding-first</td>
<td>yes</td>
<td>(F(1,65) = 11.34)</td>
<td>(p &lt; .01)</td>
<td>(\eta^2 = .15)</td>
</tr>
<tr>
<td>quiz 2</td>
<td></td>
<td>fluency-first &gt; understanding-first</td>
<td>marginally</td>
<td>(F(1,65) = 2.82)</td>
<td>(p &lt; .10)</td>
<td>(\eta^2 = .04)</td>
</tr>
<tr>
<td>Quiz: connection-understanding accuracy</td>
<td></td>
<td>condition</td>
<td>no</td>
<td>(F &lt; 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>quiz time</td>
<td>no</td>
<td>(F &lt; 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>condition * quiz time</td>
<td>no</td>
<td>(F &lt; 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quiz: connection-understanding efficiency</td>
<td></td>
<td>condition</td>
<td>no</td>
<td>(F &lt; 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>quiz time</td>
<td>no</td>
<td>(F &lt; 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>condition * quiz time</td>
<td>no</td>
<td>(F &lt; 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test: transfer accuracy</td>
<td></td>
<td>test time</td>
<td>yes</td>
<td>(F(1,66) = 3.76)</td>
<td>(p &lt; .01)</td>
<td>(\eta^2 = .12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>condition * test time</td>
<td>no</td>
<td>(F &lt; 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>posttest</td>
<td></td>
<td>understanding-first &gt; fluency-first</td>
<td>marginally</td>
<td>(F(1,65) = 3.05)</td>
<td>(p &lt; .10)</td>
<td>(\eta^2 = .05)</td>
</tr>
<tr>
<td>Test: transfer efficiency</td>
<td></td>
<td>test time</td>
<td>yes</td>
<td>(F(1,66) = 8.66)</td>
<td>(p &lt; .01)</td>
<td>(\eta^2 = .12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>condition * test time</td>
<td>no</td>
<td>(F &lt; 1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 22. Overview of results on learning outcome measures from Experiment 5.
<table>
<thead>
<tr>
<th>measure</th>
<th>problem type</th>
<th>effect</th>
<th>significant</th>
<th>F/χ²-value</th>
<th>adj. p-value</th>
<th>effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>error rate</strong></td>
<td>equivalence-sense</td>
<td>understanding-first &lt; fluency-first</td>
<td>no</td>
<td>$F(1, 66) = 2.61$</td>
<td>$p &gt; .10$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>comparison-sense</td>
<td>understanding-first ~ fluency-first</td>
<td>marginally</td>
<td>$F(1, 66) = 2.80$</td>
<td>$p &lt; .10$</td>
<td>$\eta^2 = .04$</td>
</tr>
<tr>
<td></td>
<td>equivalence-fluency</td>
<td>understanding-first &lt; fluency-first</td>
<td>marginally</td>
<td>$F(1, 66) = 2.79$</td>
<td>$p &lt; .10$</td>
<td>$\eta^2 = .04$</td>
</tr>
<tr>
<td></td>
<td>comparison-fluency</td>
<td>understanding-first ~ fluency-first</td>
<td>no</td>
<td>$F(1, 66) = 2.04$</td>
<td>$p &gt; .10$</td>
<td></td>
</tr>
<tr>
<td><strong>step duration</strong></td>
<td>equivalence-sense</td>
<td>understanding-first ~ fluency-first</td>
<td>no</td>
<td>$F &lt; 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>comparison-sense</td>
<td>understanding-first ~ fluency-first</td>
<td>no</td>
<td>$F &lt; 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>equivalence-fluency</td>
<td>understanding-first ~ fluency-first</td>
<td>no</td>
<td>$F &lt; 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>comparison-fluency</td>
<td>understanding-first ~ fluency-first</td>
<td>no</td>
<td>$F &lt; 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>representation-connections (phase 1 interviews)</strong></td>
<td>sense-making problems</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>fluency-building problems</td>
<td>understanding-first ~ fluency-first</td>
<td>no</td>
<td>$\chi^2 &lt; 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>all tutor problems</td>
<td>fluency-building problems &gt; sense-making problems</td>
<td>yes</td>
<td>$\chi^2(1, N = 187) = 88.99$</td>
<td>$p &lt; .01$</td>
<td></td>
</tr>
<tr>
<td><strong>conceptual reasoning (phase 1 interviews)</strong></td>
<td>all tutor problems</td>
<td>understanding-first &gt; fluency-first</td>
<td>yes</td>
<td>$\chi^2(1, N = 351) = 10.60$</td>
<td>$p &lt; .01$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>condition</td>
<td>no</td>
<td></td>
<td>$F(1, 21) = 1.11$</td>
<td>$p &gt; .10$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>topic</td>
<td>yes</td>
<td></td>
<td>$F(1, 21) = 11.19$</td>
<td>$p &lt; .01$</td>
<td>$\eta^2 = .35$</td>
</tr>
<tr>
<td></td>
<td>condition * topic</td>
<td>marginally</td>
<td></td>
<td>$F(1, 21) = 4.09$</td>
<td>$p &lt; .10$</td>
<td>$\eta^2 = .16$</td>
</tr>
<tr>
<td></td>
<td>equivalence</td>
<td>understanding-first &lt; fluency-first</td>
<td>marginally</td>
<td>$F(1, 21) = 3.53$</td>
<td>$p &lt; .10$</td>
<td>$\eta^2 = .14$</td>
</tr>
<tr>
<td></td>
<td>comparison</td>
<td>understanding-first ~ fluency-first</td>
<td>no</td>
<td>$F &lt; 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>frequency of switching between representations (phase 2 eye-tracking data)</strong></td>
<td>condition</td>
<td>yes</td>
<td></td>
<td>$F(1, 21) = 4.43$</td>
<td>$p &lt; .05$</td>
<td>$\eta^2 = .17$</td>
</tr>
<tr>
<td></td>
<td>topic</td>
<td>no</td>
<td></td>
<td>$F &lt; 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>condition * topic</td>
<td>yes</td>
<td></td>
<td>$F(1, 21) = 7.09$</td>
<td>$p &lt; .05$</td>
<td>$\eta^2 = .25$</td>
</tr>
<tr>
<td></td>
<td>equivalence</td>
<td>understanding-first ~ fluency-first</td>
<td>no</td>
<td>$F(1, 21) = 1.43$</td>
<td>$p &gt; .10$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>comparison</td>
<td>understanding-first &gt; fluency-first</td>
<td>yes</td>
<td>$F(1, 21) = 5.95$</td>
<td>$p &lt; .01$</td>
<td>$\eta^2 = .22$</td>
</tr>
<tr>
<td><strong>second-inspection fixations of area models (phase 2 eye-tracking data)</strong></td>
<td>condition</td>
<td>no</td>
<td></td>
<td>$F(1, 21) = 1.64$</td>
<td>$p &lt; .10$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>topic</td>
<td>yes</td>
<td></td>
<td>$F(1, 21) = 21.66$</td>
<td>$p &lt; .01$</td>
<td>$\eta^2 = .51$</td>
</tr>
<tr>
<td></td>
<td>condition * topic</td>
<td>no</td>
<td></td>
<td>$F(1, 21) = 2.46$</td>
<td>$p &lt; .10$</td>
<td></td>
</tr>
</tbody>
</table>

*Table 23.* Overview of results on process-level measures from Experiment 5.
4.5.3.1. Effects on quiz: Connection-fluency and connection-understanding

Table 24 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 (see Table 19 for an overview of the procedure).

<table>
<thead>
<tr>
<th></th>
<th>Understanding-first</th>
<th>Fluency-first</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quiz 1: connection-understanding</td>
<td>.49 (.36)</td>
<td>.39 (.38)</td>
</tr>
<tr>
<td>Quiz 2: connection-understanding</td>
<td>.41 (.33)</td>
<td>.45 (.41)</td>
</tr>
<tr>
<td><strong>Efficiency</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quiz 1: connection-understanding</td>
<td>-.04 (12.99)</td>
<td>.12 (10.05)</td>
</tr>
<tr>
<td>Quiz 2: connection-understanding</td>
<td>.41 (.33)</td>
<td>.45 (.41)</td>
</tr>
</tbody>
</table>

Table 24. Means and standard deviations (in brackets) by condition on connection-understanding and connection-fluency measures per quiz time.

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. 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, $F(1,65) = 1.42, p > .10$. Post-hoc 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, $F(1,65) = 4.52, p < .05, \eta^2 = .07$.

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, $F(1,65) = 1.17, p > .10$. Post-hoc 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. 29 shows students’ accuracy on the reproduction-fluency quizzes at quiz times 1 and 2, Fig. 30 shows students’ efficiency on the reproduction-fluency quizzes at quiz times 1 and 2.

![Graph showing accuracy and efficiency on reproduction-fluency quizzes](image)

**Fig. 29.** Accuracy on reproduction-fluency quizzes by condition by quiz time.

**Fig. 30.** Efficiency on reproduction-fluency quizzes by condition 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 ($F$s < 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 neither accuracy nor 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 25 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.

<table>
<thead>
<tr>
<th></th>
<th>Understanding-first</th>
<th>Fluency-first</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer pretest</td>
<td>.45 (.35)</td>
<td>.53 (.34)</td>
</tr>
<tr>
<td>Transfer posttest</td>
<td>.51 (.36)</td>
<td>.47 (.32)</td>
</tr>
<tr>
<td><strong>Efficiency</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer pretest</td>
<td>.34 (.44)</td>
<td>.42 (.40)</td>
</tr>
<tr>
<td>Transfer posttest</td>
<td>.51 (.48)</td>
<td>.63 (.44)</td>
</tr>
</tbody>
</table>

*Table 25.* Means and standard deviations (in brackets) 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. 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 ($F$s < 1). Post-hoc comparisons on the posttest with time spent on the Fractions Tutor as covariates shows a marginally significant advantage of the understanding-first condition at the posttest, $F(1,65) = 3.05, p < .10, \eta^2 = .05$.

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 ($F$s < 1).

Fig. 31 shows students’ accuracy on the transfer test at the pretest and the posttest. Fig. 32 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.

### 4.5.3.3. Effects on learning curves: Rates of learning

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 20) 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 equivalence-sense error rates, $F(1, 66) = 2.61, p > .10$, or on equivalence-sense error rates, $F(1, 66) = 2.04, p > .10$.

Fig. 31. Accuracy on transfer by condition by test time.

Fig. 32. Efficiency on transfer by condition by test time.
This finding provides partial support for the understanding-first errors-fluency hypothesis H1. Later in the learning sequence (that is, on comparison problems), the understanding-first condition shows lower error rates than the fluency-first condition. It is reasonable to assume that the advantage of conceptual understanding takes a while until it manifests in a decreased error rate during problem solving. Thus, when considering problem-solving behaviors later during the learning sequence, the fluency-first errors-fluency hypothesis H0 can be rejected in favor of the understanding-first errors-fluency hypothesis H1.

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. 33 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. 33 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, fluency seems to decrease students’ benefit from fluency-building problems. 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. 33. Learning curves by condition across comparison sense-making knowledge components. Bars show standard errors.](image)
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 20) as dependent measures. There were no significant differences between conditions on either dependent measure ($F$s < 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 26 shows the frequencies of representation connections and conceptual utterances obtained during phase 1 of the experiment. The interrater reliability between two independent coders was substantial with $\kappa = .66$.

<table>
<thead>
<tr>
<th></th>
<th>Understanding-first</th>
<th>Fluency-first</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation-surface</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Representation-concept-incorrect</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Representation-concept-correct</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Representation-fluency</td>
<td>76</td>
<td>82</td>
</tr>
<tr>
<td>Concept-correct</td>
<td>206</td>
<td>145</td>
</tr>
</tbody>
</table>

Table 26. Frequencies of utterances coded as representation connections and conceptual reasoning.

To test the understanding-first connections-fluency hypothesis $H1$ (that the understanding-first condition will make more connections while working on connectional fluency-building support 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 26), a chi-square test to investigate the fluency-first connections-sense hypothesis $H1$ (that the fluency-first condition will make more connections while working on connectional
sense-making support 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 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 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 only very few connection making, whereas the fluency-building problems elicited significantly more connections 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. Effects on eye-tracking measures: Visual attention

Complete eye-tracking data was available for 28 students from phase 2 of the experiment. Four further students were excluded from the analysis due to poor eye-tracking data (specifically, for these students, the average tracking ratio of lower than 70%, meaning that less than 70% of the
duration of the experiment, the eye-tracker could detect the student's eye gaze). Thus, data from \( N = 24 \) students were analyzed (\( n = 12 \) in the understanding-first condition, and \( n = 12 \) in the fluency-first condition). At this time, only the eye-gaze data collected on the sense-making problems are available for analysis. Table 27 shows the means and standard deviations on the eye-tracking variables by topic and condition.

<table>
<thead>
<tr>
<th></th>
<th>Understanding-first</th>
<th>Fluency-first</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of switching between representations</td>
<td>Equivalence 80.17 (39.31)</td>
<td>91.67 (53.73)</td>
</tr>
<tr>
<td></td>
<td>Comparison 72.42 (30.65)</td>
<td>65.17 (24.38)</td>
</tr>
<tr>
<td>Duration of second-inspection fixations of area models</td>
<td>Equivalence 1585 ms (1561 ms)</td>
<td>1911 ms (2628 ms)</td>
</tr>
<tr>
<td></td>
<td>Comparison 15281 ms (12121 ms)</td>
<td>7295 ms (4698 ms)</td>
</tr>
<tr>
<td>Duration of second-inspection fixations of number lines</td>
<td>Equivalence 48939 ms (22675 ms)</td>
<td>50565 ms (27892 ms)</td>
</tr>
<tr>
<td></td>
<td>Comparison 24884 ms (11247 ms)</td>
<td>21968 ms (15141 ms)</td>
</tr>
</tbody>
</table>

Table 27. Means (standard deviations in brackets) on eye-tracking variables.

Due to the unavailability of eye-tracking data on connectional fluency-building support, I cannot test the understanding-first eye-gaze-fluency hypothesis H1 at this time. To investigate the fluency-first eye-gaze-sense hypothesis H1 (that the fluency-first condition will exhibit more eye-gaze behaviors that are considered to indicate integrative processing and conceptual reasoning while working on the connectional sense-making support), I conducted repeated measures ANCOVAs with pretest as a covariate, and the eye-tracking variables of interest on equivalence and comparison problems as repeated, dependent variables. A repeated measures ANCOVA on frequency of switching between representations showed no main effect of condition, \( F(1, 21) = 1.11, p > .10 \), but a significant main effect of topic, \( F(1, 21) = 11.19, p < .01, \eta^2 = .35 \), and a marginally significant interaction of topic with condition, \( F(1, 21) = 4.09, p < .10, \eta^2 = .16 \). Post-hoc comparisons showed that the fluency-first condition switches marginally significantly more often between representations on equivalence problems, \( F(1, 21) = 3.53, p < .10, \eta^2 = .14 \), but not on comparison problems (\( F < 1 \)). This finding provides partial support for the fluency-first eye-gaze-sense H1 on equivalence (i.e., earlier in the learning process), but not on comparison (i.e., later in the learning process).

A repeated measures ANCOVA on duration of second-inspection fixations of area models (i.e., circle or rectangle) showed a significant main effect of condition, \( F(1, 21) = 4.43, p < .05 \),

---

8 The analysis of the eye-tracking data collected on the fluency-building problems was delayed due to difficulties with the SMI BeGaze software in defining 'moving' AOIs that move along with students' dragging-and-dropping interactions.
η² = .17, and a significant interaction of topic with condition, \( F(1, 21) = 7.09, p < .05 \), but no significant main effect of topic (\( F < 1 \)). Post-hoc comparisons showed that the understanding-first condition fixates significantly longer on area models at the second inspection than the fluency-first condition on comparison problems, \( F(1, 21) = 5.95, p < .01, \eta^2 = .22 \), but not on equivalence problems, \( F(1, 21) = 1.43, p > .10 \). This finding contradicts the fluency-first eye-gaze-sense hypothesis H1.

A repeated measures ANCOVA on duration of second-inspection fixations on number lines showed no significant main effect of condition, \( F(1, 21) = 1.64, p < .10 \), nor a significant interaction of topic with condition, \( F(1, 21) = 2.46, p < .10 \), but a significant main effect of topic, \( F(1, 21) = 21.66, p < .01, \eta^2 = .51 \). This finding does not support the fluency-first eye-gaze-sense H1.

To gain further insight into these differences in eye-gaze behaviors, I computed correlations with students' error rates on equivalence sense-making problems and comparison sense-making problems. The frequency of switching between representations on equivalence sense-making problems correlates positively with students' error rates on these problems (\( r = .358, p < .10 \); recall that a higher error rate corresponds to lower performance on the tutor problems), whereas the duration of second-inspection fixations on area models on comparison sense-making problems correlates negatively with students' error rates on these problems (\( r = -.357, p < .10 \); recall that a lower error rate corresponds to higher performance on the tutor problems).

Taken together, the only finding that is consistent with the fluency-first hypothesis is the higher frequency of switching between representations on equivalence problems. However, the frequency of switching appears to be associated with lower performance on these problems, as indicated by the positive correlation with students' error rates. Rather than indicating the integration of information across representations, students who have received connectional fluency-building support prior to working on connectional sense-making support problems may exhibit more visual confusion than students in the understanding-first condition. Yet, this interpretation remains speculative. Furthermore, the fluency-first condition show shorter duration of second-inspection fixations on area models than students in the understanding-first condition on comparison problems. Students' duration of second-inspection fixations on area models are associated with higher performance on these problems, as indicated by the negative correlation with students' error rates. This finding might indicate that receiving connectional fluency-building sup-
port before connectional sense-making support inhibits students' integration of new information with the area models presented in the worked-example part of the problems more. No differences were found on integration processes involving number lines. The lack of a difference in students' visual attention to number lines may be due to their engagement in problem-solving with the number lines, which is required by the Fractions Tutor problems.

In summary, the analysis of eye-gaze behaviors does not provide conclusive evidence for the fluency-first eye-gaze-sense H1, and therefore, the understanding-first eye-gaze-sense hypothesis H0 cannot be rejected. Yet, the interpretation of these findings remains highly speculative, and warrants further investigation. The analysis of the eye-gaze data collected on connectional fluency-building problems is likely to clarify the overall picture of these findings.

4.5.3.6. Relations between dependent measures

![Diagram](image)

**Fig. 34.** 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 “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\(^9\) 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. 34 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 on the observed covariance matrix,\(^10\) 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,\(^11\) and outputs the pattern with the best Bayesian Information Criterion (BIC) score.\(^12\) 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.

Mediation of condition effects through performance on connection-fluency quiz

First, I 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. 35 shows a model found by GES, with coefficient estimates included. The model fits the data well, \((\chi^2 = 2.43, \text{df} = 4, p = .66)\). Students

\(^9\) Tetrad, freely available at [www.phil.cmu.edu/projects/tetrad](http://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).

\(^10\) An example of a testable constraint is a vanishing partial correlation, e.g., \(\rho_{XY|Z} = 0\).

\(^11\) 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.

\(^12\) 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.
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.

Fig. 35. 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.

Next, I investigated whether the differences between conditions on efficiency on connection-fluency quiz items mediates effects on efficiency on the transfer posttest. Fig. 36 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 confirms 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 20) mediates the advantage of the understanding-first condition over the fluency-first condition on *accuracy* on the transfer posttest. Fig. 31 shows a model found by GES, with coefficient estimates included. The model fits the data well, ($\chi^2 = 4.58, df = 4, p = .33$). 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 20) 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, ($\chi^2 = 3.38, df = 3, p = .38$). 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 connectional sense-making 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 harm 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 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 which 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 into the question of how these learning processes interact. I contrasted two competing hypotheses. On the one hand, the understanding-first hypothesis posits that connectional understanding enables students to acquire connectional fluency by helping them focus on conceptually relevant aspects of graphical representations. According to the fluency-first hypothesis, on the other hand, connectional fluency frees 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 condition. 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. Yet, students in the understanding-first condition showed lower efficiency on connection-fluency quiz items, but 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. 36), students' lower efficiency on connection-fluency quiz items was associated with higher performance 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 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, whereas connectional fluency hampers students' benefit from connectional sense-making support. The mediation analysis demonstrates that these mechanisms can 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. The finding that connectional fluency harms students' benefit from connectional sense-making support, 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 show that the understanding-first condition engages in more conceptual reasoning about fractions than the fluency-first condition. At the same time, connectional fluency-building support elicits significantly more connection-making utterances than connectional sense-making support, alt-
hough 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 from making these connections, by conceptually reasoning about the connections rather than being overly influenced by the perceptual properties of the graphical representations.

Third, although the analysis of students' eye-gaze behavior remains highly speculative and inconclusive (also due to the fact that this analysis remains incomplete at this time), it seems that the results do not support the fluency-first hypothesis. It is interesting to note that the frequency of switching between graphical representations are negatively associated with students' performance on connectional sense-making problems. Rather than indicating integrative processing and connection making between graphical representations, the frequency of switching (at least on sense-making problems) might indicate superficial processing or visual confusion. Perhaps this result is in line with the finding from the retrospective interviews that the sense-making problems do not elicit much connection making altogether. Frequent switching between representations (as identified by the eye-tracking data) that does not coincide with explicit connection making (as identified by the retrospective interviews) may not be conducive to students' conceptual learning. It will be particularly interesting to investigate the association of frequency of switching on connectional fluency-building problems with measures of problem-solving performance. These observations illustrate the importance of integrating multiple measures of process-level data to disambiguate one another. If confirmed, this finding also stands in contrast to prior research on connection making between dual representations that have used the frequency of switching as measures of conceptual integration (e.g., Holsanova & Holmberg, 2009; Johnson & Mayer, 2012). Taken together, albeit inconclusive and incomplete with regard to understanding students' learning processes, the analysis of eye-tracking data provides interesting insights at the methodological level even at its current stage.

It is important to note that the advantage of the understanding-first condition over the fluency-first condition pays 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 comes 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 does not come at the expense of lower efficiency in solving fractions problems: there are no differences between the two conditions on the transfer posttest.

In summary, the results from Experiment 5 support 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. Therefore, 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.
5 Conclusion

Taken together, my dissertation work comprises a theoretical framework on learning with multiple graphical representations and a multi-representational intelligent tutoring system for fractions learning, thus reflecting two perspectives of my research: the learning sciences perspective and the educational technology perspective. At the heart of both contributions lies a sequence of controlled classroom experiments which served both to evaluate the theoretical framework and to iteratively improve the Fractions Tutor. In this section, I first review the contributions of my empirical work to the learning sciences perspective of my dissertation work, and then my contributions to the educational technology perspective of my research. After discussing limitations of my work and possible future directions, I end by reflecting on the integration of the learning sciences and educational technology perspectives of my work through the use of a multi-methods approach.
5 Conclusion

5.1 Learning sciences perspective

The learning sciences perspective of my work focuses on processes involved in learning with multiple graphical representations. Since many educational materials provide multiple graphical representations in addition to textual descriptions and symbolic representations, my work extends prior research that has mainly focused on learning with dual representations: learning with representations that use different symbol systems, such as textual descriptions accompanied with 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 & Schnitz, 2009; Suthers et al., 2008) I demonstrate that the advantage of multiple representations is not limited to learning with representations from different symbol systems and describe a set of processes involved in learning with representations that use the same symbol systems: multiple graphical representations.

A first step in my work was to investigate whether the advantage of dual representations generalizes to multiple graphical representations. Once I confirmed this notion, I turned to investigating instructional support for a number of learning processes which, based on my theoretical framework, I hypothesized to be imperative to students’ benefit from multiple graphical representations. A sequence of experimental studies confirms that these learning processes play a role in students' learning with multiple graphical representations. Furthermore, each experimental study evaluated the effectiveness of instructional support for these learning processes. Students need to be supported in learning processes that lead to representational understanding, representational fluency, connectional understanding, and connectional fluency. My research provides guidance for instructional designers on how to provide instructional support for these learning processes, thereby achieving the overarching goal of promoting robust learning of domain knowledge through the proper use of multiple graphical representations.

5.1.1 Advantage of multiple graphical 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 focus of prior research on learning with dual representations. In fact, as argued in section
2.1, existing theoretical frameworks for learning with dual representations cannot explain why multiple graphical representations might enhance learning. By contrast, they predict 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 visual information channel (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 supported in representational sense-making processes. Experiment 3 provides further support for the notion that multiple graphical representations lead to more robust domain knowledge than a single graphical representation by demonstrating that the advantage of multiple graphical representations does not depend on the specific graphical representation used in the single-representation control condition. Thus, Experiments 1 and 3 extend prior research that has focused on learning with dual representations. My research demonstrates that multiple representations of the same symbol system (e.g., multiple graphical representations) can indeed enhance learning of robust domain knowledge. In other words, multiple representations of the same symbol system (i.e., multiple graphical representations) 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 and conceptually understand the conceptual aspects of fractions instruction, 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 dual 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 of 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 are part of 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. These findings are in line with a number of studies that show that dual representations 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 processes that are prerequisite for students’ benefit from multiple graphical representations.

As I argue in the reminder of this section, my research shows that the advantage of multiple graphical representations over a single depends on students’ receiving instructional support for the learning processes described by my theoretical framework: representational sense-making processes and fluency-building processes, and connectional sense-making processes and fluency-building processes.

5.1.2 Representational sense-making processes
Experiments 1 and 2 focus on sense-making processes which lead to understanding of individual graphical representations (i.e., representational sense-making processes). While Experiment 1 investigates the effects of explicit support for representational sense-making processes by providing reflection prompts, Experiment 2 investigates the effects of implicit support for representational sense-making processes through practice schedules.

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 dual representations which show that (1) the advantage of learning with text and graphics (as opposed to text alone) may stem from dual representations enhancing students’ tendency to self-explain (Ainsworth & Loizou, 2003), and that (2) interventions that promote self-explanation activities can enhance students’ benefits from dual representations (Berthold et al., 2008; Berthold & Renkl, 2009; Zhang & Linn, 2011). Experiment 1 extends these findings from learning with dual representations to learning with multiple graphical representations (which are provided along with text and
symbols). Furthermore, Experiment 1 is, to the best of my knowledge, the first study to systematically investigate the effects of reflection prompts on students’ benefits from multiple graphical representations. The reflection prompts in the Fractions Tutor are designed to support representational sense-making processes as described in section 2.2: students relate components of each graphical representation (e.g., the number of colored sections in a circle) to the corresponding knowledge component (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 dual representations (Schwonke, Berthold et al., 2009; Seufert, 2003). Therefore, representational understanding, as described by my theoretical framework, incorporates learning processes discussed in the literature on dual representations as connection making between representations of different symbol systems. Taken together, Experiment 1 shows that students’ benefit from multiple graphical representations depends on receiving such support for representational sense-making processes.

Experiment 2 contrasts the effects of interleaving graphical representations while blocking task types and interleaving task types while blocking graphical representations on students’ conceptual understanding of graphical representations. Results show that interleaving task types while blocking graphical representations leads to higher effectiveness and efficiency in using conceptual representational knowledge to solve fractions problems than interleaving graphical representations while blocking task types. In the light of the literature on the contextual interference effect (de Croock et al., 1998; van Merriënboer et al., 2002), I argue that practice schedules affect what aspects of the learning content students abstract across (de Croock et al., 1998; Shea & Morgan, 1979) and which aspects they frequently reactivate and strengthen (de Croock et al., 1998; Lee & Magill, 1983, 1985). Interleaving task types provides students with opportunities to abstract across different applications of the same graphical representation to a sequence of different task types and thereby enhances representational sense-making processes. Abstracting across different subsequent task types in which they use the same graphical representation may help them to form an abstract understanding of the given graphical representation independent of how it is being used to solve a specific task at hand. At the same time, students frequently reactivate task-specific applications of the same graphical representation (e.g., to solve equivalence task versus addition tasks). This process of frequently retrieving knowledge about how a given
graphical representation can be used to represent a task-specific problem increases the likelihood that a student will recall that knowledge later on. Taken together, Experiment 2 extends previous research on practice schedules by demonstrating what aspect of a learning task is interleaved (i.e., graphical representation or task type) has an impact on students’ robust learning of the domain knowledge by promoting their representational understanding. Thus, representational sense-making processes are crucial to students’ robust conceptual learning from multiple graphical representations.

The finding that practice schedules have an impact on students’ representational understanding, which in turn promotes their ability to benefit from the multiplicity of graphical representations, is also interesting in the light of prior research on learning with dual 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 (Bodemér & 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 representational sense-making processes; that is, without explicit support.

It is important to note that 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. Therefore, the results from Experiment 2 by no means indicate that interleaving task types is an alternative to providing reflection prompts in order to support students’ representational understanding. Rather, in addition to providing reflection prompts, interleaving task types while blocking graphical representations can further enhance students’ benefit from multiple graphical representations by promoting representational sense-making processes.

5.1.3 Representational fluency-building processes

Experiment 3 focuses on fluency-building processes which lead to fluency in using individual graphical representations to solve domain-specific problems (i.e., representational fluency).
Building on the finding of Experiment 2 (that one should interleave task types), Experiment 3 investigates whether in addition to interleaving task types, one should also interleave graphical representations. The results show that interleaving graphical representations leads to more robust learning of domain knowledge than blocking graphical representations. Furthermore, the interleaved practice condition shows higher learning gains than students in a single-representation control condition.

I attribute this advantage to students’ gains in representational fluency; that is, their ability to fast and effortlessly use each graphical representation to solve fractions problems. By repeatedly reactivating representation-specific knowledge in the interleaved condition, students strengthen 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 in-depth analysis of different types of process-level measures supports the interpretation that interleaving graphical representations promotes representational fluency-building processes. First, an analysis of think-aloud protocols rules out an alternative explanation for the advantage of interleaving graphical representations; namely, that students spontaneously make connections between consecutively presented graphical representations. Second, Bayesian knowledge tracing analysis of the tutor log data demonstrates higher learning rates for students in the interleaved condition, compared to students in the blocked condition. This finding confirms that students in the interleaved condition become more fluent in using graphical representations to solve fractions problems than students in the blocked condition.

In applying Bayesian knowledge tracing to research on interleaved practice, Experiment 3 also extends prior research on practice schedule effects which have failed (just like myself) 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.

Taken together, Experiment 3 demonstrates that representational fluency-building processes play an important role in students’ learning of robust domain knowledge from multiple graphical
representations. Instructional designers can support this learning process by interleaving graphical representations. Experiment 3 thereby extends prior research on learning with dual representations which have mainly focused on sense-making processes (e.g., Ainsworth, 2006). Building on Koedinger and colleagues’ Knowledge-Learning and Instruction framework (2012), I argue that fluency-building processes also play a role in students’ learning with multiple graphical representations: students need to become fluent in using each individual graphical representation fast and effortlessly to solve domain-relevant problems. Although the role of fluency-building processes is considered an important factor of math learning, in the sense of developing fluency carrying out procedures and in recalling math facts (e.g., Arroyo et al., 2011), and in making connections between graphical, text-based, and symbolic math representations (Kellman et al., 2008, 2009), the importance of developing fluency with individual graphical representations has not been widely recognized. The findings from Experiment 3 demonstrate that fluency in using individual graphical representations is an important aspect of students’ learning of robust fractions knowledge. Consequently representational fluency-building process play an important role in students’ learning with multiple graphical representations.

5.1.4 Connectional sense-making and fluency-building processes
Experiments 4 and 5 focus on sense-making processes and fluency-building processes in making connections between multiple graphical representations (i.e., connectional sense-making processes and connectional fluency-building processes). While Experiment 4 establishes that connectional sense-making processes and connectional fluency-building processes interact, Experiment 5 shows that connectional fluency-building processes depend on connectional understanding.

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 us-
ing the same symbol system, whereas prior research has focused on connection making between representations using different symbol systems, such as text and graphic (Bodemer & Faust, 2006; Bodemer et al., 2004; Plötzner et al., 2001; Seufert, 2003), between symbolic representation and graphic (van der Meij & de Jong, 2006), or between text, symbolic representations, and graphic (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 students need to actively engage in connectional sense-making processes rather than relying on support that is automatically provided by the system.

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 established that dual 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 worked examples as a means to support connection making between multiple graphical representations. Experiment 4 shows that a novel application is a successful means to support students in actively engaging in connectional sense-making processes. Thereby Experiment 4 extends prior research by showing that worked examples can increase the effectiveness of multiple graphical representations more effective, when used to support connectional sense-making processes (and in conjunction with support for connectional fluency-building processes).
Taken together, Experiment 4 provides support for the prediction made by my theoretical framework that both connectional sense-making processes and fluency-building processes play a crucial role in students' benefit from multiple graphical representations. Furthermore, Experiment 4 establishes that these two learning processes are not independent of one another. Rather, they complement one another.

An analysis of the Experiment 4 log data by the means of causal path modeling provides tentative insights into how connectional sense-making processes and fluency-building processes 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 leaves 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 experience lower cognitive load while making sense of connections between graphical representations, and can therefore benefit more from connectional sense-making support.

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, conceptual reasoning, and visual attention during the learning process. Taken together, the results from Experiment 5 support 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 leads 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 higher accuracy without diminishing students’ development of efficiency in solving fractions tasks.

An analysis of process-level measures provides an explanation for this finding. On the one hand, connectional sense-making support increases students’ benefit from connectional fluency-building support by reducing errors students make on connectional fluency-building problems. Similarly, providing connectional sense-making support before connectional fluency-building support leads to increased conceptual reasoning about fractions. This finding is in line with the results from the causal path analysis in Experiment 4. On the other hand, connectional fluency-building support diminishes students’ benefit from connectional sense-making support by increasing errors students make on sense-making problems, and (although further evidence is need to support these assertions) by increasing visual confusion and reducing visual integration of new information with familiar graphical representations provided in the worked-example part of the
sense-making 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. This finding clarifies 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.

These findings stand in contrast to the assumption by Kellman and colleagues (2009, in press) that fluency in making connections can be acquired independently of connectional understanding. 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. Rather, students become more careless at solving fractions problems, evidencing their focus on becoming more efficient rather than more accurate.

The findings in Experiment 5 are consistent with Kirschner and colleagues (2006)’s critique of minimally-guided practice. Connectional fluency-building support requires students to develop perceptual expertise in a rich environment without providing conceptual guidance about how to solve the problems. As students discover which aspects of graphical representations are structurally relevant, Kirschner and colleagues (2006) would predict that they experience high cognitive load which is known to hamper learning (Chandler & Sweller, 1991). Furthermore, 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). It is important to note that Experiment 5 did not assess students’ cognitive load during the learning process, such that the interpretation of the findings in terms of cognitive load remain somewhat speculative and should be tested in future research.

Taken together, the results from Experiment 5 are quite striking if we recall that all students worked on the exact same problems, the only difference being the order in which they were provided. The sequence in which different types of support for connection making is provided influences how students work with connection-making problems, which in turn impacts what students
learn from them (i.e., efficiency in the fluency-first condition versus accuracy and efficiency in the understanding-first condition). Altogether, these findings suggest that connectional sense-making support equips students with the prerequisite knowledge to benefit from connectional fluency-building support by focusing on conceptually relevant aspects of graphical representations while acquiring connectional fluency, and to relate their knowledge about connections between graphical representations to conceptual knowledge about the domain.

5.1.5 Summary

Taken together, these experiments provide evidence that at least four learning processes play an important role in students' learning with multiple graphical representations. Whether or not students’ robust learning of domain knowledge benefits from multiple graphical representations (compared to a single graphical representation) depends on whether they receive instructional support for these learning processes. First, students need to be supported in representational sense-making processes by relating them to the abstract concept they represent (Experiment 1, see section 4.1), and by applying each graphical representation to a sequence of different task types (Experiment 2, see section 4.2). Furthermore, students need to be supported in representational fluency-building processes (Experiment 3, see section 4.3). Finally, students need to be supported in actively engaging in connectional sense-making processes, and subsequently, in connectional fluency-building processes (Experiments 4 and 5, see sections 4.4 and 4.5). The Fractions Tutor incorporates instructional support for these learning processes.
5.2 Educational technology perspective

A further contribution of my dissertation is a successful educational technology: 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 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). 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 piece
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 technology for fractions learning is in and by itself an important contribution. But in addition, my approach in integrating learning sciences research with iterative development of an educational technology based on methods originating in intelligent tutoring systems research and human-computer interaction constitutes a further contribution of my work which can benefit the development of other educational technologies 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 a theoretical framework about how students learn with multiple graphical representations. By investigating the predictions of this framework, 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. 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 representa-
tions within intelligent tutoring systems). Rather, my research provides an empirically tested theoretical framework that provides guidance for instructional design of educational technologies that use multiple graphical representations.

My research not only shows which learning processes play a role in learning with multiple graphical representations, but also how best to support these learning processes through instructional design. By investigating the effectiveness of different types of instructional support for learning processes involved in learning with multiple graphical representations, my research provides directions for which types of support are effective in helping students engage in these crucial learning processes. 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 with the learning processes they support. For instance, the use of Bayesian knowledge tracing in Experiment 3 (see section 4.3.3) provides evidence that interleaving graphical representations supports representational fluency-building processes, while think-aloud protocols ruled out the possibility that interleaving graphical representations supports connectional sense-making processes (see section 4.3.2). The use of causal path analysis modeling in Experiments 4 and 5 (see sections 4.4.3.2 and 4.5.3.6) establishes 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. For example, knowing 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), 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 the human-computer interaction perspective which has substantially contributed to the success of the Fractions Tutor within the context of real educational settings. The most important contribution of the human-computer interactions perspective to the design of the Fractions Tutor consists in the way in which I resolved 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 which 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-method 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 Merrienboer 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) arbitrary, 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 evaluated 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 which 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 create a goal hierarchy for different types of users, in conducting 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 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 which 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. I expect that wherever multiple graphical representations are employed with this purpose, students need to engage in representational sense-making and fluency-building processes as well as in connectional sense-making and fluency-building processes. Yet, this assertion is somewhat speculative and 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 dual representations), to learning with multiple representations from the same symbol system (i.e., multiple graphical representations). My theoretical framework provides guidance for such research in that it makes specific predictions 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 consistent-
ly been provided in an interleaved fashion across all conditions, based on the findings of Experiment 2 (see section 4.2)? Again, my research does not answer that question. More broadly speaking, it remains an open question whether support for 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, my research was conducted within a very specific type of educational technology: intelligent tutoring systems. 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. An interesting direction for future research would be to investigate the effects of providing instructional support for the 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, provide 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 outcome and process-level measures, my research yields insights that motivate 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 which were experimentally contrasted in Experiment 5. The results from Experiment 5, in turn, result 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) an empirically validated theoretical framework, (2) an effective intelligent tutoring system, (3) a set of instructional design principles that can be integrated in a wide range of educational technologies, and (4) a new methodology that can guide instructional designers in resolving conflicts that inevitably arise in the context of real educational settings.
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 Tune-Pekkan, Laurens Feenstra, Cathy Kramer, Mitch Nathan, Peg Smith, Cassandra Studer, Jenny Olson, 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, and Mike Komisin.

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 at Barrett Elementary, Dickson Elementary, Donaldson Elementary, Edgewood Elementary, Fairless Elementary, Francis McClure Intermediate School, Hampton Middle School, Harrison Park Elementary, Holy Trinity Catholic School, Hopewell Elementary, Independence Elementary, Level Green Elementary, Margaret Ross Elementary, McCullough Elementary, McKee Elementary, Park Elementary School, Shaffer Elementary, Steel Valley Middle School, Sunrise Elementary, Trafford Elementary, Wilkins Elementary, Wilson Elementary, and Woodland Hills Academy.

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


Leikin, R., Leikin, M., Waisman, I., & Shaul, S. (2013). Effect of the presence of external representations on accuracy and reaction time in solving mathematical double-choice
problems by students of different levels of instruction. *International Journal of Science and Mathematics Education*.


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.


Appendix 1: Pennsylvania State Standards

<table>
<thead>
<tr>
<th>Pennsylvania State Standards</th>
<th>Alignment of Fractions Tutor with the Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use whole numbers and fractions to represent quantities. (2.1.3. B) Use drawings, diagrams</td>
<td>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.</td>
</tr>
<tr>
<td>or models to show the concept of fraction as part of a whole. (2.1.3. D)</td>
<td></td>
</tr>
<tr>
<td>Use models to represent fractions and decimals (2.1.5. D)</td>
<td>Circles, rectangles, and number lines are used throughout the Fractions Tutor units</td>
</tr>
<tr>
<td>Develop and apply algorithms to solve word problems that involve addition, subtraction, and/or</td>
<td>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.</td>
</tr>
<tr>
<td>multiplication with fractions and mixed numbers that include like and unlike denominators.</td>
<td></td>
</tr>
<tr>
<td>(2.2.5. C)</td>
<td></td>
</tr>
</tbody>
</table>

Table A1. Alignment of the Fractions Tutor with Pennsylvania State Standards.
## Appendix 2: NCTM Standards

<table>
<thead>
<tr>
<th>NCTM Standards</th>
<th>Alignment of Fractions Tutor with the Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td>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).</td>
</tr>
<tr>
<td>All students should use models, benchmarks, and equivalent forms to judge the size of fractions.</td>
<td>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.</td>
</tr>
<tr>
<td>All students should explore numbers less than 0 by extending the number line and through familiar applications.</td>
<td>The Fractions Tutor provides practice in interpreting and manipulating fractions using circles, rectangles, and number lines.</td>
</tr>
</tbody>
</table>

*Table A2. Alignment of the Fractions Tutor with Pennsylvania State Standards.*
## Appendix 3: Common Core Standards

<table>
<thead>
<tr>
<th>Grade</th>
<th>Common Core Standards</th>
<th>Alignment of Fractions Tutor with the Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Understand a fraction $\frac{1}{b}$ as the quantity formed by 1 part when a whole is partitioned into $b$ equal parts; understand a fraction $\frac{a}{b}$ as the quantity formed by a parts of size $\frac{1}{b}$. (3.NF.1.)</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>Understand a fraction as a number on the number line; represent fractions on a number line diagram. (3.NF.2.)</td>
<td>The fractions tutor includes activities with the number line while supporting students in making connections between circles, rectangles, and number lines.</td>
</tr>
<tr>
<td></td>
<td>Represent a fraction $\frac{1}{b}$ 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 $\frac{1}{b}$ and that the endpoint of the part based at 0 locates the number $\frac{1}{b}$ on the number line. Represent a fraction $\frac{a}{b}$ on a number line diagram by marking off a lengths $\frac{1}{b}$ from 0. Recognize that the resulting interval has size $\frac{a}{b}$ and that its endpoint locates the number $\frac{a}{b}$ on the number line.</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>Explain equivalence of fractions in special cases, and compare fractions by reasoning about their size. (3.NF.3.)</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>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., $\frac{1}{2} = \frac{2}{4}$, $\frac{4}{6} = \frac{2}{3}$. 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 $&gt;$, $=$, or $&lt;$, and justify the conclusions, e.g., by using a visual fraction model.</td>
<td></td>
</tr>
</tbody>
</table>
### Table A3. Alignment of the Fractions Tutor with Common Core Standards.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Description</th>
<th>Fractions Tutor Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.NF.1.</td>
<td>Explain why a fraction ( \frac{a}{b} ) is equivalent to a fraction ( \frac{(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.</td>
<td>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.</td>
</tr>
<tr>
<td>4.NF.2.</td>
<td>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 &gt;, =, or &lt;, and justify the conclusions, e.g., by using a visual fraction model.</td>
<td>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.</td>
</tr>
<tr>
<td>4.NF.3.</td>
<td>Understand a fraction ( \frac{a}{b} ) with ( a &gt; 1 ) as a sum of fractions ( \frac{1}{b} ). 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.</td>
<td>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.</td>
</tr>
<tr>
<td>5.NF.1.</td>
<td>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.</td>
<td>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.</td>
</tr>
</tbody>
</table>
Appendix 4: Experiment 2 test items

Fig. A4-1. Example test items from the representational knowledge test used in Experiment 2.
Fig. A4-2. Example test items from the operational knowledge test used in Experiment 2.
Appendix 5: Experiment 3 test items

Fig. A5-1. Example test items from the area models fluency test used in Experiment 3.

Fig. A5-2. Example test items from the number lines fluency test used in Experiment 3.
Fig. A5-3. Example test items from the conceptual transfer 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?

A  

B  

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. A5-4. Example test items from the procedural 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.)

$0\begin{array}{ccc} 0 & \frac{1}{4} & \frac{2}{4} \end{array}$

Compare $\frac{3}{6}$ and $\frac{11}{12}$. Which fraction is larger? Explain your thinking with pictures or words.

Compare $\frac{1}{7}$ and $\frac{5}{8}$. Which fraction is larger? Explain your thinking with pictures or words.
Appendices

Appendix 6: Experiment 4 test items

2) Place a dot on the number line that shows \( \frac{8}{5} \).

3) Which of the number lines below correctly shows \( \frac{1}{4} \)?
   a) \( \frac{1}{4} \)
   b) \( \frac{1}{4} \)
   c) \( \frac{1}{4} \)
   d) \( \frac{1}{4} \)

7) There is a bag of candies. If 10 candies are \( \frac{2}{3} \) of all the candies in the bag, how many total candies are in the bag?
   There are \( \square \) candies in the bag.

4) Which of the number lines below correctly shows \( \frac{4}{5} \)?
   a) \( \frac{4}{5} \)
   b) \( \frac{4}{5} \)
   c) \( \frac{4}{5} \)
   d) \( \frac{4}{5} \)

12) Students in Mrs. Johnson’s class were asked to tell why \( \frac{4}{5} \) is greater than \( \frac{2}{3} \).
    Whose reason is best?
    a) Kelly said, ‘Because \( 4 \) is greater than \( 2 \).’
    b) Ken said, ‘Because \( 5 \) is larger than \( 3 \).’
    c) Kim said, ‘Because \( \frac{4}{5} \) is closer than \( \frac{2}{3} \) to 1.’
    d) Kevin said, ‘Because \( 4 \times 5 \) is more than \( 2 \times 3 \).’

6) The blue rectangle below shows \( \frac{3}{4} \). Which gray rectangle shows the unit?
   a) \( \frac{3}{4} \)
   b) \( \frac{3}{4} \)
   c) \( \frac{3}{4} \)
   d) \( \frac{3}{4} \)

5) The blue circle below shows \( \frac{2}{3} \). Which gray figure shows the unit?
   a) \( \frac{2}{3} \)
   b) \( \frac{2}{3} \)
   c) \( \frac{2}{3} \)
   d) \( \frac{2}{3} \)

13) Mark says \( \frac{1}{4} \) of his candy bar is smaller than \( \frac{1}{5} \) of the same candy bar.
    Is Mark right?
    a) Yes
    b) No

Use words to explain why you think Mark is right or wrong.

Fig. A6-1. Example test items from the conceptual knowledge test used in Experiment 4.
Fig. A6-2. Example test items from the procedural knowledge test used in Experiment 4.
Appendix 7: Experiment 5 test items

Fig. A7. Example test items from the transfer test used in Experiment 5.