Interleaved practice in multi-dimensional learning tasks: Which dimension should we interleave?

Martina A. Rau [a,*], Vincent Aleven [a], Nikol Rummel [b]

[a] Human-Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
[b] Institute of Educational Research, Ruhr-Universität Bochum, Germany

ARTICLE INFO

ABSTRACT

Research shows that multiple representations can enhance student learning. Many curricula use multiple representations across multiple task types. The temporal sequence of representations and task types is likely to impact student learning. Research on contextual interference shows that interleaving learning tasks leads to better learning results than blocked practice, but this research has not investigated whether it matters on which dimension we interleave learning tasks. Many educational materials include multiple task types and multiple representations. Should we interleave representations or task types? We conducted a classroom experiment to investigate the effects of interleaving task types (while blocking representations) and interleaving representations (while blocking task types). The participants (158 5th- and 6th-graders) worked with a corresponding version of an intelligent tutoring system for fractions. Our results show an advantage for interleaving task types over interleaving representations. These results extend prior work on contextual interference by showing that this effect is sensitive to the dimension being interleaved. We also extend the literature on learning with multiple representations by investigating the effect of interleaved practice with different representations. The results provide guidance to designers of complex curricula.

Published by Elsevier Ltd.

1. Introduction

Graphical representations are universal educational tools: flow diagrams are used in programming, schemas and tree diagrams in biology, charts and graphs in mathematics — to mention just a few examples. Realistic learning materials usually include learning tasks that vary on several dimensions: they use multiple graphical representations over a sequence of multiple task types. The goal in using multiple graphical representations is to promote robust knowledge that can be transferred to unfamiliar tasks and that lasts over time (see Koedinger, Corbett, & Perfetti, in press) by supporting conceptual understanding of the different representations, and by enhancing procedural knowledge covered across multiple task types.

In areas where learners engage in extended problem-solving practice, instructors and instructional designers must decide how to sequence representations (for example, in the domain of fractions, circle diagrams, and number lines) and task types (for example, finding equivalent fractions, comparing fractions, and adding fractions). Should they interleave graphical representations

while blocking task types, or should they interleave task types while blocking representations? What sequence will lead to the most robust learning gains? The decision of how to sequence task types and representations is likely to influence learners' acquisition of robust conceptual and procedural knowledge.

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

In this paper, we describe a classroom experiment that compares the effect of interleaving graphical representations to the effect of interleaving task types on students' robust learning of fractions. Specifically, we conducted an experiment that contrasts

the effect of interleaving multiple graphical representations while blocking task types to the effect of interleaving task types while blocking multiple graphical representations. We assessed students' robust learning of fractions with regard to the effectiveness and efficiency of conceptual knowledge of fractions representations and their procedural knowledge with fractions operations. To ensure the ecological validity of our results, the experiment was carried out in classrooms in the context of a proven educational technology, namely, Cognitive Tutors (Koedinger & Aleven, 2007; Koedinger & Corbett, 2006). Cognitive Tutors pose rich problem-solving tasks and allow the use of interactive graphical representations. Cognitive Tutors are used in a large number of classrooms across the United States (Corbett, Koedinger, & Hadley, 2001; Koedinger & Corbett, 2006) and therefore represent a realistic scenario for mathematics instruction. Specifically, we used a web-based tutoring system that covers a wide range of topics for early fractions learning. The system uses a number of interactive graphical representations of fractions (i.e., the commonly used number line, circle, and set).

Before describing the experimental study in detail, we discuss relevant literature on learning with multiple representations and on learning with graphical representations of fractions, as well as the research on the contextual interference effect.

1.1. Learning with multiple representations

Well-designed representations are powerful learning tools because they emphasize crucial conceptual aspects of the learning material (Ainsworth, 2006). Further, skillful and flexible use of representations (e.g., in the context of problem solving) is considered a key aspect of expertise in complex domains (Goldman, 2003), including mathematics (Common Core State Standards Initiative, 2010; Kilpatrick, Swafford, & Findell, 2001; National Council of Teachers of Mathematics, 2000, 2006; National Mathematics Advisory Board Panel, 2008) and science (Kozma, 2003; National Research Council, 2002; National Science Teachers Association, 2009).

Research in a variety of domains has demonstrated that multiple representations have the potential to substantially promote learning (e.g., Ainsworth, Bibby, & Wood, 1998; Eliam & Poyas, 2008; Lewalter, 2003; Rasch & Schnotz, 2009; Schnotz & Bannert, 2003; Seufert, 2003). However, the majority of the research on learning with multiple representations has investigated the benefits of pairing text and graphics used for a single task type (e.g., Ainsworth, Bibby, & Wood, 2002; Bodemer & Faust, 2006; Bodemer, Ploetzner, Feuerlein, & Spada, 2004; Butcher, 2006; Eliam & Poyas, 2008; Lewalter, 2003; Rasch & Schnotz, 2009; Seufert, 2003). Furthermore, in most of the research on learning with multiple representations, the different representations are provided together, for instance, text and graphic side-by-side. Many realistic curricula (especially in mathematics) are more complex: they integrate multiple graphical representations that are used interchangeably across multiple task types over an extended period of time. Often, the graphical representations are sequenced across learning tasks (i.e., only one graphical representation is provided per learning task, but different graphical representations are provided across multiple learning tasks). Typical fractions instruction, for example, includes a variety of graphical representations, each of which conveys a slightly different conceptual interpretation of fractions. For instance, circles depict fractions as parts of a whole, number lines depict fractions as measures, and sets depict fractions as ratios (Charalambous & Pitta-Pantazi, 2007; Cramer, Behr, Post, & Lesh, 1997a, 1997b; Lamon, 2005). Our own informal review of commonly used U.S. elementary and middle-school mathematics curricula for fractions (Bennett, 2004; Fitzgerald, Lappan, & Fey,

2004; Hake, 2004; Kilpatrick et al., 2001; Lappan, Fey, & Fitzgerald, 1998) showed that these curricula employ a rather wide variety of graphical representations, and that each of these graphical representations is commonly used across a variety of task types.

Several studies have found benefits for learning with graphical representations of fractions. For example, Mack (1995) reports on a case study that showed that graphical representations such as circles can help students overcome misconceptions. The representations help students relate their burgeoning formal knowledge of fractions to their existing informal knowledge of sharing and dividing. The Rational Number Project curriculum uses a variety of representations, including circles, number lines, chips, and symbols (Cramer, Wyberg, & Leavitt, 2009). In a large-scale experimental study, this curriculum was shown to significantly improve students' understanding of fractions, compared to a standard curriculum (Cramer, Post, & delMas, 2002). A curriculum that emphasizes various linear representations of fractions, developed by Moss and Case (1999), was shown to be more effective than a standard curriculum for the fractions topics that were covered. Caldwell (1995) argues, based on a case study with 5th- and 6th-graders, that area models of fractions are useful learning tools. Yang and Reys (2001) describe the use of number lines in their classrooms to illustrate the importance of using graphical representations of fractions in helping students understand rational number concepts. Finally, in our own experimental study, we found that students working with a version of an intelligent tutoring system that uses multiple graphical representations of fractions learn more deeply than students who work with a system version that presents only a single graphical representation, although only when prompted to reflect on how the graphical representations of fractions (e.g., half of a circle) relate to the symbolic representation (e.g., 1/2) (Rau, Aleven, & Rummel, 2009). Although the studies on learning with multiple graphical representations of fractions have typically employed graphical representations across a variety of task types, they have not systematically investigated the effects of different practice schedules of graphical representations.

Taken together, research both from educational psychology and mathematics education support the notion that learning with multiple (graphical) representations can help students' learning. However, this research has not investigated the effects on students' learning of interleaving or blocking representations.

1.2. Research on the contextual interference effect

The literature on contextual interference gives reason to believe that the decision of whether to interleave graphical representations or task types is likely to influence students' robust learning of fractions. Generally, the results from contextual interference research have demonstrated that "interleaved practice" leads to better learning results than "blocked practice" (Battig, 1972; Schmidt & Bjork, 1992). In this research, the independent variable has typically been whether learning tasks were presented in "blocks" of the same type (e.g., task 1 − task 1 − task 1 − task 2 − task 2 − task 2 − task 3 − task 3 − task 3), or whether learning tasks of different types were interleaved (e.g., task 1 − task 2 − task 3 − task 1 − task 2 − task 3 − task 1 − task 2 − task 3). The contextual interference effect has been demonstrated in a variety of domains including vocabulary learning (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Pashler, Rohrer, Cepeda, & Carpenter, 2007), motor tasks (Hebert, Landin, & Solmon, 1996; Li & Wright, 2000; Meiran, 1996; Meiran, Chorev, & Sapir, 2000; Schmidt & Bjork, 1992; Schneider, 1985; Simon & Bjork, 2001), and cognitively more complex tasks such as solving algebra problems in mathematics (Rohrer & Taylor, 2007; Taylor &

Rohrer, 2010), troubleshooting (de Croock, Van Merriënboer, & Paas, 1998; Van Merriënboer, Schuurman, de Croock, & Paas, 2002), and in decision-making tasks (Helsdingen, van Gog, & Van Merrienboer, 2011). However, this research has not investigated whether the dimension on which the learning tasks are interleaved (e.g., representation or task type) matter. In other words, it remains an open question whether problem solving with multiple representations is more effective when representations are interleaved or when dissimilar task types are interleaved.

Two explanations have been offered for the common finding that interleaving learning tasks leads to more robust learning than blocking them. First, it has been argued that when working on interleaved learning tasks, learners have to *reactivate* the knowledge needed to solve each learning task more often than when working on blocked learning tasks (De Croock, Van Merrienboer, & Paas, 1998; Lee & Magill, 1983, 1985; Sweller, 1990). Reactivation is likely to occur more frequently with interleaved task sequences than with blocked task sequences. In blocked task sequences, the knowledge used in one learning task can often be used in the next task. Some or all of it can presumably be kept in working memory across task boundaries, and does not have to be reactivated. In interleaved task sequences, by contrast, the required knowledge more often needs to be retrieved from long-term memory. Cognitive theories such as ACT-R (Anderson, 1993, 2002) hold that repeated retrieval strengthens associations between cues and associated elements in long-term memory. Reactivating knowledge by retrieving it from long-term memory therefore increases the likelihood that this knowledge can be recalled later on.

An alternative explanation for the advantages of interleaving is that it helps students to *abstract* knowledge from different learning tasks presented consecutively (De Croock et al., 1998; Shea & Morgan, 1979). Abstraction requires that important knowledge from the previous task is still in working memory when the new task is "loaded" into working memory (e.g. by reading the task instructions). When knowledge related to different tasks is active in working memory, the student can compare the knowledge relevant to the respective learning tasks, consciously or unconsciously. By comparing across learning tasks, learners may see more clearly which task properties are key and which are incidental, thereby directing learners' attention to processes relevant to knowledge construction (Bannert, 2002; Paas & Van Gog, 2006; Van Merriënboer et al., 2002) and helping them abstract the knowledge common to consecutive learning tasks. Encountering dissimilar learning tasks back-to-back occurs frequently in interleaved task sequences, but infrequently in blocked task sequences. We note that the two explanations are not mutually exclusive and that in a realistic learning sequence both learning mechanisms may be active. It may be, however, that situations that are most conducive to reactivation may not be most conducive to abstraction, as discussed further below.

1.3. Research questions

Generally, in any domain that uses multiple graphical representations across multiple task types, students should acquire conceptual understanding of the representations used, that is, the ability to interpret and construct the representations and to use representations to make sense of domain-relevant problems. For instance, in the case of fractions, students should be able to use a circle diagram to answer the question: "What fraction of the pizza is left?" We use the term "representational knowledge" to denote this kind of knowledge. Furthermore, students should acquire procedural knowledge of the operations covered by the learning material, such as finding equivalent fractions or comparing

fractions, both with and without representations ("operational knowledge").

The goal of our classroom experiment is to investigate the impact of interleaving learning tasks on two different dimensions with regard to students' acquisition of robust knowledge of fractions, which includes both representational and operational knowledge. In particular, should learners frequently switch between task types while switching infrequently between representations (e.g., A1 − A2 − A3 − B1 − B2 − B3 − C1 − C2 − C3, where A, B, and C are different representations, and the numbers stand for three different task types)? Or should learners frequently switch between representations while infrequently switching between task types (e.g., A1 − B1 − C1 − A2 − B2 − C2 − A3 − B3 − C3)? We know of no studies that have addressed this question, yet as we have argued, it is of both practical and theoretical importance.

How, based on the theoretical accounts for the contextual interference effect described above, might one expect interleaving representations or task types to affect students' acquisition of robust knowledge of fractions? It is difficult to make a prediction, because (1) there are substantial differences in individual students' developmental trajectories, and (2) robust knowledge of fractions is a complex construct. Based on the theoretical accounts of contextual interference just discussed, however, we hypothesize that *it will be most effective to interleave learning tasks along the dimension with the greatest variability,* that is, the dimension along which the learning tasks vary the most from task to task. We hypothesize that the task types (e.g., equivalent fractions, fraction comparison, fraction addition, etc.) are more variable than the graphical representations we use in our instructional material (i.e., circles, sets, and number lines).

Arguably, the different task types used in our experiment are more saliently different than multiple representations for several reasons. First, the different task types require students to apply different operations (such as finding equivalent fractions, or adding fractions). By contrast, the different representations provide different conceptual views on the task at hand (by depicting a fraction as a shaded part of a circle, or as a dot on a number line), and the conceptual differences (i.e., fractions as parts of a whole, or fractions as a measure) might be difficult to discern for novice learners. Second, the graphical representations are designed to be intuitive: graphical representations typically employ perceptual processes in an effective and easy-to-understand way. They may also be intuitive in the sense that they connect to informal prior knowledge, such as, in the domain of fractions, everyday notions of sharing and dividing equally. In order to use graphical representations, students are not expected to engage in explicit reasoning about the properties of the representations. The different task types covered in our experiment, on the other hand, require students explicitly to use different procedures to solve the task. Due to these properties of graphical representations and task types, we expect that the conceptual differences between them will not be as salient as the differences between task types. For this reason, we anticipate that task types are the more variable dimension, compared to graphical representations.

We expect that interleaving tasks types on the more variable dimension will give students better opportunities for reactivation and (to a certain degree) for abstraction. Students are expected to reactivate (and thereby strengthen) any knowledge that is not shared between consecutive learning tasks. We expect this reactivation process to happen more frequently when learning tasks differ on the more variable dimension. The more dissimilar consecutive learning tasks are, the greater the need for reactivation, and (consequently) the higher the chance that interleaved practice will increase students' acquisition of robust knowledge. Therefore,

interleaving learning tasks on the most variable dimension may be most conducive to reactivation.

Abstraction may be more likely to occur when learning tasks are interleaved on a moderately variable dimension. Consecutive tasks need to be sufficiently dissimilar so that students can compare the knowledge associated with the different tasks; without dissimilarities, there is nothing to abstract across. On the other hand, it is crucial for abstraction to occur that the consecutive learning tasks share some common knowledge that can be abstracted from them. If the learning tasks are too dissimilar, the knowledge required to solve them may not be similar enough for students to see the correspondences to abstract from the learning tasks, so that abstraction cannot occur. In other words, if abstraction is the mechanism by which the interleaving of learning tasks leads to better learning, higher variability of consecutive learning tasks should not always lead to higher learning gains. Rather, there might be an optimal level of variability in learning tasks, such that consecutive learning tasks are similar enough to allow for abstraction, but dissimilar enough so that learners are likely to abstract knowledge from them.

Given that the arguments presented above are not specific to representational knowledge or operational knowledge, we expect that interleaving task types will have a stronger effect than interleaving representations both on representational knowledge and on operational knowledge.

The present classroom experiment aims at investigating the effect of blocking graphical representations versus interleaving graphical representations on students' robust knowledge of fractions. We assess robust knowledge of fractions both with regard to the *effectiveness* and *efficiency* of students' knowledge. Both are educationally relevant measures. Effectiveness measures solely take into account students' increase in performance over time, as measured by their score on repeated administration of similar test forms (e.g., average scores across multiple test items). However, performance alone is an incomplete measure of students' learning: many tests also include time constraints, in part because students are expected to become faster at performing tasks that they have mastered (Koedinger et al., in press). We expect that, as they learn, students become more efficient at solving a task: the acquisition of robust knowledge should be reflected not only in performing better, but also in performing faster. We predict that interleaving of task types while blocking graphical representations (int-types), as compared to interleaving of graphical representations while blocking task types (int-reps), will lead to: more effective representational knowledge (hypothesis 1a); more efficient representational knowledge (hypothesis 1b); more effective operational knowledge (hypothesis 2a); more efficient operational knowledge (hypothesis 2b).

2. Methods

2.1. Participants

The study involved 158 students in grades 5 and 6, aged 9–12 years, from 16 classes of a total of three schools. Students participated in the study during their regular mathematics instruction. All schools were located in the greater Pittsburgh area.

2.2. Materials

2.2.1. Fractions Tutor

Students worked with different versions of a web-based intelligent tutoring system for fractions designed and created specifically for this study. The Fractions Tutor is a type of Cognitive Tutor, as mentioned. 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. In a variety of research studies, Cognitive Tutor courses for mathematics have been shown to lead to substantial achievement gains in comparison with traditional classroom instruction (Anderson, Corbett, Koedinger, & Pelletier, 1995; Corbett et al., 2001; Koedinger, 2002; Ritter, Anderson, Koedinger, & Corbett, 2007; Ritter, Kulikowich, Lei, McGuire, & Morgan, 2007). At the heart of the Fractions Tutor lies a cognitive model of students' problem-solving steps. The model captures skill components in the form of production rules, following the ACT-R (and other) theories of cognition.
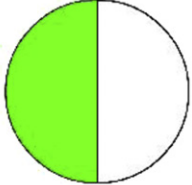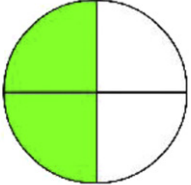
In our research, we make use of a newer version of this intelligent tutor technology, called example-tracing tutors (Aleven, McLaren, Sewall, & Koedinger, 2009). Example-tracing tutors are behaviorally similar to Cognitive Tutors, meaning that they provide step-by-step guidance in the form of feedback and on-demand hints. In contrast to Cognitive Tutors, example-tracing tutors rely on generalized examples of correct and incorrect solution paths rather than on a rule-based cognitive model of student behavior. We created the Fractions Tutor with the Cognitive Tutor Authoring Tools (CTAT; Aleven et al., 2009), designing tutor interfaces separately for each problem type and representation. The design of the interfaces and of the interactions students engage in during problem-solving are based on a number of small-scale user studies that we conducted in our laboratory, on a prior classroom experiment (Rau et al., 2009), as well as on Cognitive Task Analysis of the learning domain (Baker, Corbett, & Koedinger, 2007; Clark, Feldon, Van Merrienboer, Yates, & Early, 2008). Furthermore, an experienced mathematics teacher was involved in developing the tutor problems. All graphical representations were interactive, virtual manipulatives (Moyer, Bolyard, & Spikell, 2002). Research has demonstrated that students can benefit from using virtual manipulatives of fractions (Reimer & Moyer, 2005), and that virtual manipulatives can be at least as effective in supporting students' learning as physical manipulatives (Suh, Moyer, & Heo, 2005). The Fractions Tutor is available to students and teachers on a free website (https://mathtutor.web.cmu.edu).

The Fractions Tutor used in the study included three different *graphical representations of fractions*: circles (see Fig. 1), number lines (see Fig. 2), and sets (see Fig. 3). Each graphical representation emphasizes certain aspects of the different 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 pieces. The number line represents a measurement representation. This representation emphasizes that fractions can be compared in terms of their magnitude and that they fall between whole numbers. Finally, the set is a ratio representation and presents fractions in the context of a comparison between two quantities that are depicted as discrete objects. The *task types* covered by the Fractions Tutor cover basic concepts (e.g., identifying fractions from graphical representations and understanding fractions in terms of sharing activities) and fraction operations (e.g., equivalent fractions, comparing fractions, and adding fractions). Table 1 summarizes the task types covered by the Fractions Tutor.

In each tutor problem, the students first solved the problem by manipulating the graphical representation and then solved the problem symbolically. For instance, students could partition the interactive graphical representations into sections (for the number line), pieces (for the circle), or objects (for the sets). Fig. 1 shows an example of an equivalent fractions problem with the circle. Fig. 2 shows a corresponding problem with the number line, and Fig. 3 with the set representation. Problems were introduced with a realistic cover story. The use of cover stories is common practice in
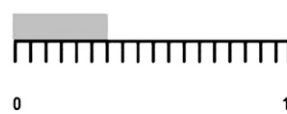
Fig. 1. Example of an equivalent fractions problem with the circle.

United States mathematics instruction at the elementary, middle-school, and high-school levels. Prior research demonstrates that providing cover stories that emphasize the deep structure of the problem can promote transfer (e.g., Baranes, Perry, & Stigler, 1989; Carraher, Carraher, & Schliemann, 1987; Thußbas, 2001). Cover stories may also facilitate understanding of the problem at hand and invoke informal strategies (Koedinger & Nathan, 2004). We chose cover stories that matched the graphical representation and that emphasized the conceptual properties of each graphical representation. For instance, the circle diagram was introduced to students as a cake or a pizza, whereas number line problems used lengths and distances as examples. The tutor provided support that

is typical of intelligent tutoring systems (VanLehn, 2006). It provided correctness feedback on all steps, indicating whether the step is right or wrong. It was able to recognize common student errors and provide error-specific feedback messages. These messages were designed to make students reconsider their answer either by reminding them of a previously-introduced principle or by providing them with an explanation for their error. Furthermore, students could request hints on all problem-solving steps. Hint messages usually had three levels. First, students received a clari-fication of the goal (e.g., "You now added all pieces into the same circle. Before you know what fraction of the whole cake you won, you need to divide the circle into equally sized pieces."). They were

Fig. 2. Example of equivalent fractions problem with the number line.

Fig. 3. Example of equivalent fractions problem with sets.

Table 1
Task types and concepts covered by Fractions Tutor.

Number	Task type	Example
1	Identify fractions from graphical representations	What fraction does the circle show?
2	Identify equivalent fractions from graphical representations	You are looking for circles that show 1/4. Select all of the below circles that apply.
3	Identify embedded fractions from graphical representations	What fraction of the colored sections in the circle is blue?
4	Identify equivalent embedded fractions from graphical representations	You are looking for circles in which 1/4 of all colored sections are blue. Select all of the below circles that apply.
5	Fractions as proper sharing of quantities	You want to share 3 pizzas among 4 people. What fraction does each of you get?
6	Fractions as quotative sharing	You have 3 pizzas that you share with your friends. Each of you gets 3/4. How many people shared the pizza?
7	Making equivalent fractions with graphical representations	The circle shows 1/4. Can you show the same fraction in eighths?
8	Recognizing equivalent fractions with graphical representations	The circle shows 1/4. Which of the circles below show an equivalent fraction?
9	Comparing fractions	You have 1/4 of a pizza, and your friend has 1/3 of a pizza. Who has more pizza?
10	Determining the magnitude of differences between fractions	You have 1/4 of a pizza, and your friend has 1/3 of a pizza. How much more does he have?
11	Adding fractions with sum smaller than 1	You have 1/4 of a pizza pepperoni and 1/3 of a pizza margarita. How much pizza do you have altogether?
12	Adding fractions with sum larger than 1	You have 3/4 of a pizza pepperoni and 2/3 of a pizza margarita. How much pizza do you have altogether?

then given conceptually oriented help, by reminding them of a specific concept (e.g., "The pieces are part of the same cake. Therefore, you keep the same denominator in the sum fraction."). Finally, students received explicit instructions regarding the next step (e.g., "Please divide the circle into four pieces.").

At the end of each problem, the tutor provided students with reflection questions in order to prompt them to make sense of their problem solution. In a prior classroom experiment, we found these prompts to be effective when students learn with multiple graphical representations of fractions (Rau et al., 2009). Students selected their answer from a drop-down menu, as shown in Fig. 2. Previous research has shown that even simple ways of prompting for self-explanations in an intelligent tutoring system can promote a self-explanation effect and lead to more robust learning (Aleven & Koedinger, 2002).

2.2.2. Test instruments

To assess students' robust knowledge of fractions, we 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). We used a structural equation model to conduct the confirmatory factor analysis for our tests. The fit of the model was determined using a variety of fit statistics, following the recommendations by Arbuckle and Wothke (1999), and by Kline (1998). We used RMSEA to determine the absolute model fit, NFI and TLI to determine the incremental fit of the model, and CMIN/df to determine our model's parsimony fit. A model is considered to have a good fit when the RMSEA is below 0.05, an NFI and TLI above 0.9, and a CMIN/df below 1.5. The fit statistics for our model were RMSEA = 0.29, NFI = 0.951, TLI = 0.982, and CMIN/df = 1.303. The test's theoretical structure was replicated with the pretest data from the current experiment.

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). Our 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., in press). Figs. 4 and 5 show 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. We 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 the 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. We 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. We 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.

Each response on each test item was scored as either correct or incorrect. When students had to type in a fraction, students received partial credit for each correct numerator and denominator, respectively. Non-reduced fractions were graded as correct. The

representational knowledge scale included a total of seven test items. The operational knowledge scale included a total of five test items. In order to make the scores more interpretable, we report relative scores (i.e., scores out of a maximum of 1).

We 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, we used a measure of efficiency described by Van Gog and Paas (2008) and by Lewis and Barron (2009). Specifically, we 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)

We followed Van Gog and Paas (2008) and Lewis and Barron (2009) and applied the concept of condition efficiency (Paas and 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. We used the time students spent on the test rather than the time they spent with the tutoring system for two reasons. First, we were 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
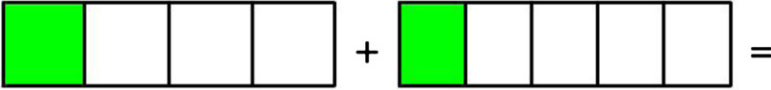
Fig. 4. Sample item of the representational knowledge scale of the fractions test. Students have to use an unfamiliar representation to identify the length of the vehicles as the crucial aspect of the representation, and relate the two lengths to form a fraction.
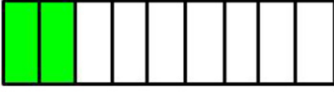
Fig. 5. Sample item of the operational knowledge scale of the fractions test. Students have to add two fractions as represented with an unfamiliar graphical representation and select their answer from a list of options.

with the tutoring system during their regular mathematics periods, which are, due to their nature, restricted in time.

2.2.3. Tutor logs

All student interactions with the tutoring system were recorded by the system. The tutor logs capture all student actions and all tutor responses. Thus, the logs include information on the number of attempts students need to solve a step correctly, and how many errors they make per step. As is common in Cognitive Tutor research, we identified the knowledge components that students need to solve each problem, based on cognitive task analysis. Knowledge components are units of cognitive function needed to solve a set of structurally similar tasks (Koedinger et al., in press). Table 2 provides examples of knowledge components we identified and for each, a description of the steps in fractions problems in which that knowledge component is required. As described further below, these knowledge components also served as a basis for the analysis of the tutor log data. Steps in tutor problems were mapped to knowledge components.

2.3. Research design

The goal of this study was to systematically investigate the effects of interleaving task types (int-types) versus interleaving representations (int-reps). We assigned students randomly 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 a single graphical representation. Each problem also involved the symbolic representation of fractions and a problem statement in text, but this was kept constant across conditions. Fig. 6 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; see Table 1). 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.[1] 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. 6), the task types are maximally interleaved and the representations are maximally blocked. That is, students covered all twelve fraction task types with one graphical representation before switching to the next representation, again working through all task types before switching to the third graphical representation (corresponding to 36 problems per representation). In this condition, students encountered a new task type after every single problem. By contrast, in the int-reps condition (see the table on the right in Fig. 6), 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, we implemented different plausible orders of graphical representations as a control factor to counterbalance potential ordering effects. Students never

---

[1] 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.

Table 2
Examples of knowledge components used in the Fractions Tutor.

Task types	Steps within tasks	Knowledge component
1. Identify fractions from graphical representations	Determine the denominator of given graphical representation	Knowing how a graphical representation depicts the denominator
1. Identify fractions from graphical representations	Determine the numerator of given graphical representation	Knowing how a graphical representation depicts the numerator
2. Identify equivalent fractions from graphical representations	Multiple choice to identify the graphical representation showing the given fraction	Being able to match equivalent graphical representations
5. Fractions as proper sharing of quantities	Determine the denominator of fractional amount per person when sharing a quantity	Knowing how to identify the denominator of the fractional part of a shared quantity
5. Fractions as proper sharing of quantities	Determine the numerator of fractional amount per person when sharing a quantity	Knowing how to identify the numerator of the fractional part of a shared quantity
6. Fractions as quotative sharing	Typing in the number of people sharing an amount given the fractional amount per person	Knowing how to identify the number of people sharing a quantity given the fractional amount per person
6. Fractions as quotative sharing	Typing in the number of wholes shared given the fractional amount per person	Knowing how to identify the number of wholes shared given the fractional amount per person
7. Making equivalent fractions with graphical representations	Multiplication factor of denominator	Knowing how to expand a fraction by multiplying numerator and denominator by the same number
7. Making equivalent fractions with graphical representations	Multiplication factor of numerator	Knowing how to expand a fraction by multiplying numerator and denominator by the same number
9. Comparing fractions	Determine the common denominator of two fractions in order to compare them	Knowing how to find the common denominator of two given fractions in order to compare them
9. Comparing fractions	Determine the numerator of a given fraction by expanding it in order to compare it to another	Knowing how to expand the numerator of a fraction in order to compare it to another
9. Comparing fractions	Multiple choice to identify the graphical representation showing the given fraction	Being able to match a graphical representation to a given fraction
9. Comparing fractions	Select the relationship (larger, smaller, equal) between two given fractions	Being able to judge the relative relationship between two fractions that have a common denominator
11. Adding fractions	Determine the denominator of sum fraction	Knowing how to find the denominator of the result from adding two fractions that have a common denominator
11 + 12. Adding fractions	Determine the numerator of sum fraction	Knowing how to add the numerators of two fractions that have a common denominator
11 + 12. Adding fractions	Partitioning the graphical representation into a total number of sections corresponding to the least common denominator	Knowing into how many sections to partition a representation to match a given symbolic fraction

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). We randomly assigned students 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. 6 thus reflects only one of the implemented orders.

2.4. Procedure

The presented study took place at the end of the school year 2008/2009. Students' regular mathematics 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.

We assessed students' knowledge of fractions three times. On the first day, students completed a pretest. They then worked on the Fractions Tutor, for 5 h, 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., in press), we gave students an equivalent delayed posttest. Students could take as much time as they needed to complete the tests. We asked the participating teachers not to revisit fractions between the immediate and the delayed posttest.

3. Results

Table 3 provides the means and standard deviations for the effectiveness of representational and operational knowledge, for time-on-task on the representational and operational knowledge
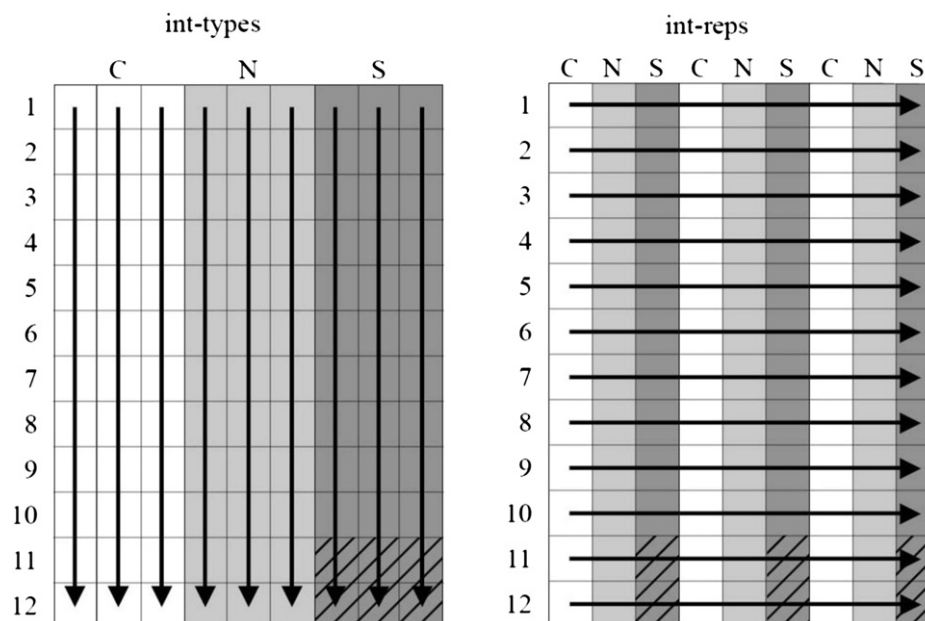
Fig. 6. Rows depict twelve different topics covered, columns shows one of several possible orders of representations: C/white = circle, N/light-gray = numberline, S/dark-gray = set. The dashed cells indicate task-type/representation combinations that were not implemented in the tutor.

subscales of the test, and for representational efficiency and operational efficiency.

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 $F(3, 97) = 1.21$, $p > .10$,

we disregarded the order of representation in the reported analyses. Since students had seen the same test that they received at the delayed posttest either at the pretest or at the immediate posttest, we 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, we 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 (see Raudenbush & Bryk, 2002) with four nested levels was used to analyze the data. At level 1, we modeled performance for each of the two posttests for each student. At level 2, we accounted for differences between students. At level 3, we modeled differences between classes, and at level 4, we accounted for differences between schools. In addition, we used post-hoc comparisons to clarify the effect of blocking versus interleaving. More specifically, the following hierarchical linear model was fitted to the data: $score_{ij} = condition_i + test\text{-}time_j + condition_i*test\text{-}time_{ij} + pre\text{-}score_i + tutor\text{-}exposure_i + school_i + class_i(school)$

$$score_{ij} = condition_i + test - time_j + condition_i*test - time_{ij}$$
$$+ pre - score_i + tutor - exposure_i + school_i$$
$$+ class_i(school)$$

$$(2)$$

$Score_{ij}$ is a student's score on the representational knowledge or the operational knowledge subscale of the fractions test (either effectiveness or efficiency score); $condition_i$ indicates whether the student was in the int-types or the int-reps condition; $test\text{-}time_j$ is an indicator variable for either the immediate or the delayed posttest; $condition_i*test\text{-}time_{ij}$ captures an interaction between condition and test-time; $pre\text{-}score_i$ is a students' score on representational knowledge or operational knowledge at the pretest (either effectiveness or efficiency score); $tutor\text{-}exposure_i$ indicates

Table 3
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.

		Pretest	Immediate posttest	Delayed posttest
Representational	int-types	0.55 (0.20)	0.61 (0.27)	0.59 (0.24)
effectiveness	int-reps	0.53 (0.22)	0.52 (0.26)	0.39 (0.32)
Operational	int-types	0.38 (0.31)	0.51 (0.34)	0.44 (0.36)
effectiveness	int-reps	0.43 (0.33)	0.40 (0.35)	0.39 (0.30)
Representational	int-types	100.18 (29.53)	70.89 (22.46)	66.95 (23.60)
time-on-task	int-reps	103.63 (32.19)	74.88 (28.33)	79.24 (30.52)
Operational	int-types	68.74 (31.64)	47.69 (18.53)	43.88 (19.09)
time-on-task	int-reps	69.77 (37.01)	46.11 (21.31)	42.71 (22.22)
Representational	int-types	−0.35 (0.89)	0.48 (0.83)	0.52 (0.80)
efficiency	int-reps	−0.48 (0.98)	0.16 (0.90)	−0.31 (1.01)
Operational	int-types	−0.50 (0.90)	0.31 (0.83)	0.24 (0.77)
efficiency	int-reps	−0.42 (1.12)	0.12 (0.78)	0.17 (0.63)

Table 4
Overview of study results on differences between conditions on representational effectiveness and representational efficiency.

measure	Test-time	Main effects/ interaction effects	Tendency of pairwise comparisons	Significant (yes/no)	F/t-value	Adj. p-value	Effect size
Representational effectiveness		Condition		Yes	$F(1, 100) = 23.97$	$p < .01$	Partial $\eta^2 = .11$
		Test-time		Yes	$F(1, 100) = 4.99$	$p < .05$	Partial $\eta^2 = .01$
		Condition*test-time		Yes	$F(1, 100) = 7.32$	$p < .01$	Partial $\eta^2 = .05$
	Immediate posttest		int-types > int-reps	Yes	$t(100) = 2.34$	$p < .05$	$d = .37$
	Delayed posttest		int-types > int-reps	Yes	$t(100) = 5.55$	$p < .01$	$d = .88$
Representational efficiency		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	$t(100) = 4.74$	$p < .01$	$d = .21$

how many problems a student solved on the relevant knowledge components being tested by either the representational knowledge test or the operational knowledge test (see Table 2 for examples of the knowledge components covered by the tutoring system); $school_i$ is a random intercept for the students' school; and $class_i(school)$ is a nested random-intercept for the student's class.

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

3.1. Differences between conditions

To investigate hypothesis 1a, that the int-types condition will outperform the int-reps condition on effectiveness of representational knowledge, we applied the HLM in Equation (2) to students' effectiveness scores on the representational knowledge subscale of the test. Table 4 provides an overview of the learning results on both the representational effectiveness and representational efficiency measures. Table 5 summarizes the least squared means and standard deviations generated by the model for representational effectiveness and representational efficiency. We found 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, we found 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 support hypothesis 1a: the int-types condition outperforms the int-reps condition on effectiveness of representational knowledge.

Table 5
Least squared means and standard deviations (in parentheses) for representational effectiveness and representational efficiency at immediate posttest, delayed posttest by condition.

		Immediate posttest	Delayed posttest
Representational effectiveness	int-types	0.58 (0.05)	0.57 (0.05)
	int-reps	0.50 (0.05)	0.36 (0.05)
Representational efficiency	int-types	0.41 (0.22)	0.45 (0.22)
	int-reps	0.03 (0.22)	−0.43 (0.22)

To investigate hypothesis 1b, that the int-types condition will outperform the int-reps condition on effectiveness of representational knowledge, we applied the HLM in Equation (2) to students' efficiency scores on the representational knowledge subscale of the test. We 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, we found an advantage for int-types over int-reps on the immediate posttest, $t(100) = 2.34$, $p < .05$, $d = .37$, and the delayed posttest, $t(100) = 5.55$, $p < .01$, $d = .88$. Our findings thus support hypothesis 1b: int-types condition outperforms the int-reps condition on effectiveness of representational knowledge.

To investigate hypothesis 2a, that the int-types condition will outperform the int-reps condition on effectiveness of operational knowledge, we applied the HLM in Equation (2) to students' effectiveness scores on the operational knowledge subscale of the test. Table 6 provides an overview of the learning results on both the operational effectiveness and operational efficiency measures. Table 7 summarizes the least squared means and standard deviations generated by the model for operational effectiveness and operational efficiency. On operational effectiveness, we 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 $\eta^2 < .01$. We found no significant interaction effect for operational effectiveness, $F(1,100) = 1.36$, $p > .10$. These results do not support hypothesis 2a: the int-types condition does not outperform the int-reps condition on effectiveness of operational knowledge.

To investigate hypothesis 2b, that the int-types condition will outperform the int-reps condition on efficiency of operational knowledge, we applied the HLM in Equation (2) to students' efficiency scores on the operational knowledge subscale of the test. On operational efficiency, we found no significant main effect of condition, $F < 1$. The effect of test-time was not significant for operational efficiency, $F < 1$. We found no significant interaction effect for operational efficiency, $F < 1$. Taken together, we cannot support hypothesis 2b: the int-types condition does not outperform the int-reps condition on efficiency of operational knowledge.

3.2. Analysis of tutor logs

To explore the effects of the int-types and the int-reps conditions on students' learning, we analyzed the tutor log data.

Table 6
Overview of study results on differences between conditions, obtained from HLM described in Equation (2).

test scale	Test-time	Main effects/ interaction effects	Tendency of pairwise comparisons	Significant (yes/no)	F/t-value	Adj. p-value	Effect size
Operational effectiveness		Condition		No	$F(1,100) = 2.05$	$p > .10$	
		Test-time		Yes	$F(1,100) = 3.04$	$p < .10$	Partial $\eta^2 < .01$
		Condition*test-time		No	$F(1,100) = 1.36$	$p > .10$	
	Immediate posttest		int-types > int-reps	No	$t < 1$		
	Delayed posttest		int-types > int-reps	No	$t < 1$		
Operational efficiency		Condition		No	$F < 1$		
		Test-time		No	$F < 1$		
		Condition*test-time		No	$F < 1$		
	Immediate posttest		int-types > int-reps	No	$t < 1$		
	Delayed posttest		int-types > int-reps	No	$t < 1$		

Specifically, we examined "learning curves" using the DataShop web service (Anderson, 1993; Koedinger et al., 2010; VanLehn et al., 2007), which depict the average error rate (across students and knowledge components) as a function of the amount of prior practice (i.e., the number of opportunities a student has had to apply a given knowledge component). Following standard practice in Cognitive Tutors research, we viewed each step in a tutor problem as a learning opportunity for the particular knowledge component involved in the step. As mentioned, as part of our cognitive task analysis during tutor development, we had identified fine-grained knowledge components (see Table 2 for examples) such that each step in a tutor problem could be mapped to one of these components. Further, we considered a step in a tutor problem to be correct if the student solved it without hints and 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). We expect that, if learning occurs, error rates will decrease with the number of learning opportunities students have encountered. Fig. 7 shows the aggregate learning curves based on error rates across knowledge components for the int-types condition and the int-reps condition. The slope of the learning curves decreases equally for both conditions. Therefore, the log data provide evidence for gradual learning in both conditions, but they do not provide evidence that the int-types and the int-reps conditions had different effects on students' learning rates.

4. Discussion and conclusions

The goal of our classroom experiment was to investigate whether interleaving task types will lead to more robust representational and operational knowledge of fractions than interleaving graphical representations. Taken together, our results provide an affirmative answer with regard to representational knowledge, but not with respect to operational knowledge. As hypothesized, our results show that interleaving task types while blocking graphical representations leads to both higher effectiveness and higher efficiency in answering questions that require representational knowledge, compared to interleaving graphical representations while blocking task types (hypothesis 1a and

Table 7
Least squared means and standard deviations (in parentheses) for operational effectiveness and operational efficiency at immediate posttest, delayed posttest by condition.

		Immediate posttest	Delayed posttest
Operational effectiveness	int-types	0.51 (0.04)	0.43 (0.04)
	int-reps	0.41 (0.04)	0.39 (0.04)
Operational efficiency	int-types	0.25 (0.11)	0.18 (0.11)
	int-reps	0.17 (0.10)	0.23 (0.10)

hypothesis 1b, respectively). Overall, our results provide support for the notion that interleaving task types leads to more robust representational knowledge, as compared to interleaving representations. On the other hand, the practice schedule of task types and graphical representations did not significantly affect students' effectiveness or efficiency in answering questions that require operational knowledge (hypothesis 2a and hypothesis 2b, respectively).

How might we explain the differences between conditions on representational knowledge (hypotheses 1a and 1b)? We have argued that interleaving learning tasks along the dimension of greatest variability is most effective, and we have argued that the differences among task types are more salient than the differences among representations. Although the different representations used in our study emphasize conceptually different aspects of representations, students might not perceive these dissimilarities because the representations are designed to be intuitive and easy to interpret. For this reason, it may be difficult for the relative novice students in our study to discern the conceptual differences between the different representations. Greater problem variability is likely to increase the need for repeated reactivation of knowledge, which leads to strengthening of that knowledge and increases the likelihood of long-term retention of that knowledge. Greater variability may also, to a degree, increase the number of opportunities for abstraction. If tasks are very similar, it may be difficult for students to abstract: there must be a sufficient number of dissimilarities to abstract over; otherwise, there is no "grist" for the "abstraction mill". It is possible that the subtle differences between representations make it difficult for students to abstract across them, at least without explicit support from the learning environment. In fact, research in other domains has demonstrated that relating representations is a difficult task (e.g., Bodemer et al., 2004; De Jong et al., 1998; Van der Meij & de Jong, 2006). Instead, interleaving task types may encourage students to abstract across different applications of the same graphical representations, which appears to lead to a more robust understanding of graphical representations than abstracting across different 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, our results suggest that the reactivation and abstraction across task types is more beneficial to students' conceptual understanding of representations than abstracting across representations.

Whether the notion that, in domains in which learning tasks vary across multiple dimensions (e.g., representations and task types), one should interleave learning tasks along the dimension with the greatest variability (i.e., in the present experiment task
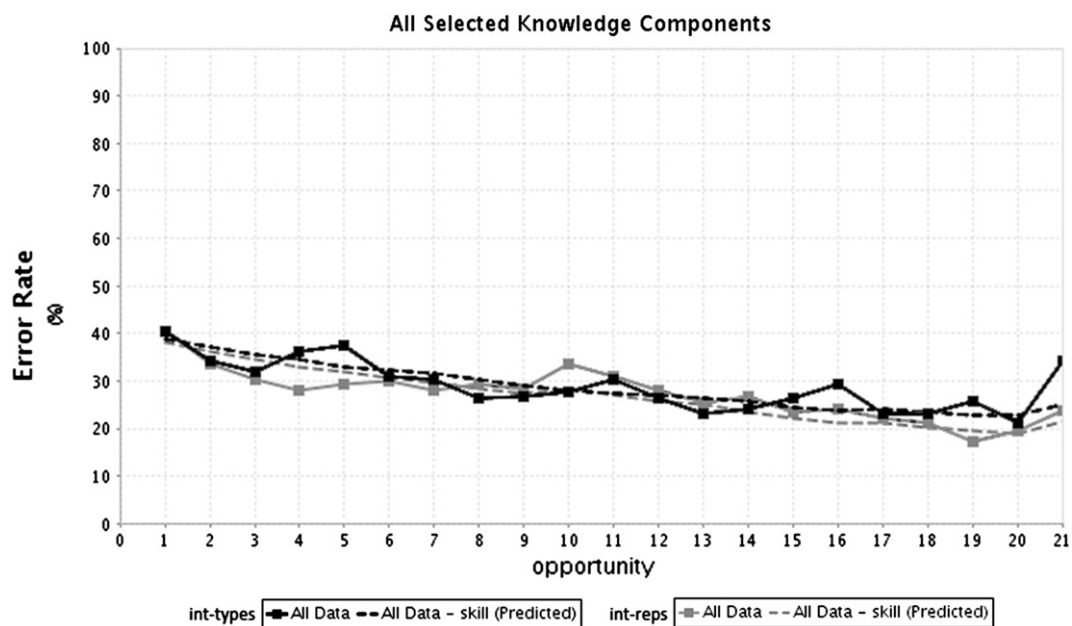
Fig. 7. Overall learning curve for the int-types condition (black) and the int-reps condition (gray). The solid lines indicate the empirical error-rate values per learning opportunity as obtained from the tutor log data. The dashed lines indicates the fitted learning curve computed from an Additive Factor Model that uses an Item-Response Model to predict how a student will perform on each knowledge component on each learning opportunity.

types and not representations) holds as a general principle should be addressed in future research. Such research should investigate the effects of interleaving different dimensions of complex learning tasks, where the dimensions differ with regard to their variability. There are few studies that have demonstrated benefits for inter-leaving learning tasks in complex domains (De Croock et al., 1998; Helsdingen et al., 2011; van Merrienboer et al., 2002; Rohrer & Taylor, 2007; Taylor & Rohrer, 2010), but none of them have investigated whether the variability of the dimension on which learning tasks are interleaved matter. As argued, interleaving highly dissimilar learning tasks may decrease the likelihood that abstraction can occur. However, in situations where consecutive learning tasks are so dissimilar that they do not require sufficiently similar knowledge to abstract across, we would not predict an advantage of interleaved practice through abstraction. In other words, if abstraction is the prominent mechanism, we would expect that a moderate level of dimension variability will lead to the best learning results. On the other hand, if reactivation is the more prominent mechanism of interleaved practice, one may expect that even interleaving maximally dissimilar learning tasks will enhance students' robust learning. Therefore, whether one should *always* interleave the dimension will likely depend on the mechanisms underlying the effect of interleaved practice, and on the nature of the dimension that is being interleaved. More research is needed to investigate boundary cases of interleaved practice to establish whether increasing the variability of the dimension on which learning tasks are being interleaved ceases to be beneficial after reaching a certain degree of dissimilarity. Our study presents a first attempt at answering this important theoretical and practical question, and we think that it has the potential to stimulate fruitful future research.

The fact that interleaving task types versus interleaving graphical representations affects only representational knowledge (hypotheses 1a and 1b) but not, as we had predicted, operational knowledge (hypotheses 2a and 2b) may reflect differences in these knowledge types. The representational knowledge scale requires primarily conceptual knowledge about how to interpret

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

Our findings also raise some additional open questions regarding the effect of interleaved practice on students' performance during practice. A common finding in the contextual

interference literature is that interleaved practice schedules facilitate long-term retention and performance on transfer tests, but lead to lower performance during the acquisition phase than blocked practice schedules (e.g., De Croock et al., 1998; Helsdingen et al., 2011; Lee & Simon, 2004; Shea & Morgan, 1979; Wulf & Shea, 2002). In our study, however, we found no difference across conditions in the performance during practice, as indicated by our analysis of the tutor log data, described above (see Fig. 7). Our study differs from prior work in that we did not compare a blocked condition versus an interleaved condition. Rather, we contrasted two different interleaved practice schedules that interleaved different aspects of the learning task. Both kinds of interleaved schedules may have created some level of interference during the acquisition phase, such that we would not necessarily expect to see a difference between the conditions with respect to students' performance during their work on the tutoring system. We note that research on practice schedules in other domains has not always found that interleaved practice leads to reduced performance during the acquisition phase (Hebert et al., 1996; Helsdingen et al., 2011; Immink & Wright, 1998; Lee & Magill, 1985; Nelson, 2006; Ollis, Button, & Fairweather, 2005; Wrisberg & Liu, 1991). A common explanation for the lack of differences during the acquisition phase is that the amount of interference created in the interleaved conditions was not sufficient (e.g., Hebert et al., 1996; Nelson, 2006). Their reasoning is similar to our argument – that the interference created by our two interleaved conditions was not sufficiently different to show an effect during students' practice with the tutoring system.

Immink and Wright (1998), and Helsdingen et al. (2011) offer a different explanation for their lack of differences on students' performance during the practice phase. Based on Immink and Wright's (1998) finding that the advantage of performance during training in a blocked practice condition relative to an interleaved practice condition disappeared when learners were given more time, they propose that performance differences during the acquisition phase are not apparent in self-paced learning settings (Helsdingen et al., 2011), because learners can plan their next steps before executing them (Immink & Wright, 1998). Sufficient time to solve a learning task might increase students' cognitive resources to encounter the increase in processing demands due to interleaved practice. Having time to plan the next steps may therefore buffer the negative effect of interleaved practice on students' performance during the acquisition phase. Future research should address the question under which circumstances we can expect effects of interleaved practice during the acquisition phase, and how the presence or absence of differences during practice relates to differences between practice schedules on posttest performance.

It is interesting to reflect on whether the results of the current study generalize beyond some of the specific properties of the learning environment we employed, an intelligent tutoring system for fractions with interactive graphical representations. Are our findings specific to the task types and graphical representations used in instructional materials for fractions learning? We expect not. Many domains employ graphical representations across multiple task types. In most of these domains, graphical representations are designed so that learners can interpret them easily, by making use of intuitive perceptual processes and by building on students' tacit prior knowledge of the real world (for example, linear graphs in coordinate systems build on of students' intuitive understanding of growth). In many of these domains, the conceptual differences between graphical representations are also less explicitly stated than the differences between task types, because task types might require the use of different procedures, whereas representations rely on perceptual properties. Consequently, we expect that interleaving of task types will lead to more robust

learning than interleaving representations also in other domains than fractions. Since learning with multiple graphical representations has broad application in a variety of domains, including mathematics, biology, statistics, economy, and many more, the potential range of applications of our findings is large.

We also expect our results to generalize to other formats than intelligent tutoring systems, such as paper-based curricula. A crucial difference between computer-based learning environments and paper-based learning materials is that the latter tend not to involve virtual manipulatives such as the interactive graphical representations used in this classroom experiment. We do not expect the results to change in favor of interleaving representations when static representations are used. Virtual manipulatives are more engaging than static representations (e.g., Moyer et al., 2002; Rogers, 1999), and therefore might motivate students more than static representations to deeply process the conceptual properties of the graphical representations. Therefore, we would expect the difference between static representations to be even less salient than between the virtual manipulatives used in the present study. In any case, task types should remain the more variable dimension, compared to representations.

Students' motivation may also play a role in the effect of interleaving task types. It is possible that interleaving task types might be perceived as less repetitive than interleaving representations, because task types are more saliently different than representations. The advantage of interleaving task types over interleaving representations might be mediated by these potential effects on students' motivation. Higher motivation may have lead students to process the material covered by the Fractions Tutor more deeply, which may have accounted for the higher learning gains in the int-types condition. Future research could address this question by assessing students' motivation during the acquisition phase and analyze whether the effects of interleaving task types versus representations is in part or fully mediated through students' motivation.

When designing complex and realistic learning materials that use multiple representations, there are many other possible ways to sequence representations and task types. Our classroom experiment contrasted interleaving task types while blocking representations, and interleaving representations while blocking task types. These two conditions represent two extremes, where other options are possible. For example, one could interleave along neither dimension (i.e., have "blocks" of practice problems that all involve the same task type and representation), or one could interleave along both dimensions. Neither of these approaches seems like a good idea: not interleaving means one misses out on the learning benefits of interleaving, whereas extreme interleaving along two dimensions may lead to cognitive overload. It is possible, however, to highly interleave one aspect while also moderately interleaving the other aspect. Whether interleaving multiple representations is helpful in addition to interleaving task types remains an open question that future research should address.

A further open question regards the mechanisms underlying the differences we found between conditions. As argued before, the mechanisms of reactivation and abstraction may not be mutually exclusive: both mechanisms may be active at the same time. While we cannot definitively conclude from our data which learning mechanism accounts for the advantage of interleaving task types, we speculate that reactivation may be the more frequent mechanism. Reactivation is a process that occurs automatically when students have to consecutively use knowledge to solve a task that is currently not active in working memory and needs to be "loaded" again into working memory. Abstraction, on the other hand, is more likely to require conscious and effortful comparison across task types. Successfully comparing across task types would require

learners to see what task properties are key to solving each problem, and which are incidental. This is a complicated process that would induce a substantial amount cognitive load, on top of the cognitive load due to successfully solving the problem at hand. We think that it is unlikely that students spontaneously engage in such a demanding comparison across different task types. However, further research is needed to identify the mechanisms underlying the advantage of interleaved practice in complex educational settings.

In conclusion, the results from our classroom experiment are both of practical and of theoretical significance. Our results also provide guidance for developers of learning materials that include multiple graphical representations which are used across multiple task types. We recommend that they interleave task types and block representations in order to promote conceptual understanding of the representations. Furthermore, from a theoretical perspective, we provide evidence that the dimension on which learning tasks are interleaved matters; interleaving task types influences the acquisition of robust conceptual knowledge more than interleaving of graphical representations does. We extend the literature on learning with multiple representations by demonstrating that the temporal sequence of multiple representations has an effect on students' robust conceptual understanding of representations. As a general principle, to be confirmed by future research, it appears that interleaving along the dimension with greatest variability is most effective.

Acknowledgments

References

Ainsworth, S. (2006). DeFT: a conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183–198. http://dx.doi.org/10.1016/j.learninstruc.2006.03.001.

Ainsworth, S. E., Bibby, P. A., & Wood, D. J. (1998). Analysing the costs and benefits of multi-representational learning environments. In M. W. van Someren, P. Reimann, H. P. A. Boshuizen, & T. de Jong (Eds.), *Learning with multiple representations*. Pergamon: Oxford.

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences, 11*(1), 25–61. http://dx.doi.org/10.1207/S15327809JLS1101_2.

Aleven, V., & Koedinger, K. R. (2002). An effective meta-cognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*(2), 147–179. http://dx.doi.org/10.1207/s15516709cog2602_1.

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

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

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

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: lessons learned. *Journal of the Learning Sciences, 4*(2), 167–207. http://dx.doi.org/10.1207/s15327809jls0402_2.

Arbuckle, J. L., & Wothke. (1999). *AMOS 4.0 User's Guide (S. 395–416)*. Chicago: Small Waters Corp.

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

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign-language vocabulary and the spacing effect. *Psychological Science, 4*(5), 316–321. http://dx.doi.org/10.1111/j.1467-9280.1993.tb00571.x.

Bannert, M. (2002). Managing cognitive load − recent trends in cognitive load theory. *Learning and Instruction, 12*(1), 139–146. http://dx.doi.org/10.1016/S0959-4752(01)00021-4.

Baranes, R., Perry, M., & Stigler, J. (1989). Activation of real-world knowledge in the solution of word problems. *Cognition and Instruction, 6*, 287–318. http://dx.doi.org/10.1207/s1532690xci0604_1.

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

Bennett, J. M. (2004). *Holt middle school math*. Holt Rinehart & Winston.

Bodemer, D., & Faust, U. (2006). External and mental referencing of multiple representations. *Computers in Human Behavior, 22*(1), 27–42. http://dx.doi.org/10.1016/j.chb.2005.01.005.

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction, 14*(3), 325–341. http://dx.doi.org/10.1016/j.learninstruc.2004.06.006.

Butcher, K. R. (2006). Learning from text with diagrams: promoting mental model development and inference generation. *Journal of Educational Psychology, 98*(1), 182–197. http://dx.doi.org/10.1037/0022-0663.98.1.182.

Caldwell, J. H. (1995). Communicating about fractions with pattern blocks. *Teaching Children Mathematics, 1*(9), 156–161.

Carraher, T., Carraher, D., & Schliemann, A. (1987). Written and oral mathematics. *Journal for Research in Mathematics Education, 18*(2), 83–97. http://dx.doi.org/10.2307/749244.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354–380. http://dx.doi.org/10.1037/0033-2909.132.3.354.

Charalambous, C. Y., & Pitta-Pantazi, D. (2007). Drawing on a theoretical model to study students' understandings of fractions. *Educational Studies in Mathematics, 64*(3), 293–316. http://dx.doi.org/10.1007/s10649-006-9036-2.

Clark, R. E., Feldon, D. F., Van Merrienboer, J. J. G., Yates, K. A., & Early, S. (2008). Cognitive task analysis. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 577–594). Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Common Core State Standards Initiative. (2010). *Common core state standards (Mathematics standards)*. Retrieved 07.04.11 from. http://www.corestandards.org/the-standards/mathematics.

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

Cramer, K., Behr, M., Post, R., & Lesh, R. (1997a). *Rational number project: Fraction lessons for the middle grades: Level 1*. Dubuque, IA: Kendall/Hunt Publishing.

Cramer, K., Behr, M., Post, R., & Lesh, R. (1997b). *Rational number project: Fraction lessons for the middle grades: Level 2*. Dubuque, IA: Kendall/Hunt Publishing.

Cramer, K. A., Post, T. R., & delMas, R. C. (2002). Initial fraction learning by fourth- and fifth-grade students: a comparison of the effects of using commercial curricula with the effects of using the rational number project curriculum. *Journal for Research in Mathematics Education, 33*(2), 111–144. http://dx.doi.org/10.2307/749646.

Cramer, K., Wyberg, T., & Leavitt, S. (2009). *Rational number project*. Fraction Operations & Initial Decimal Ideas.

De Croock, M. B. M., Van Merrienboer, J. J. G., & Paas, F. (1998). High versus low contextual interference in simulation-based training of troubleshooting skills: effects on transfer performance and invested mental effort. *Computers in Human Behavior, 14*(2), 249–267. http://dx.doi.org/10.1016/S0747-5632(98)00005-3.

De Jong, T., Ainsworth, S. E., Dobson, M., Van der Meij, J., Levonen, J., Reimann, P., et al. (1998). Acquiring knowledge in science and mathematics: the use of multiple representations in technology-based learning environments. In M. W. Van Someren, W. Reimers, H. P. A. Boshuizen, & T. de Jong (Eds.), *Learning with multiple representations*, (Oxford).

Eliam, B., & Poyas, Y. (2008). Learning with multiple representations: extending multimedia learning beyond the lab. *Learning and Instruction, 18*(4), 368–378. http://dx.doi.org/10.1016/j.learninstruc.2007.07.003.

Fitzgerald, F., Lappan, P., & Fey, J. T. (2004). *Connected mathematics (number and operations grade 6, bits and pieces I, understanding rational numbers)*. Pearson Prentice Hall.

French, K. E., Rink, J. E., & Werner, P. H. (1990). Effects of contextual interference on retention of three volleyball skills. *Perceptual and Motor Skills, 71*, 179–186.

Goldman, S. R. (2003). Learning in complex domains: when and why do multiple representations help? *Learning and Instruction, 13*(2), 239–244. http://dx.doi.org/10.1016/S0959-4752(02)00023-3.

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

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

Helsdingen, A., van Gog, T., & Van Merrienboer, J. J. G. (2011). The effects of practice schedule and critical thinking prompts on learning and transfer of a complex judgement task. *Journal of Educational Psychology, 103*(2), 383–398. http://dx.doi.org/10.1037/a0022370.

Immink, M. A., & Wright, D. L. (1998). Contextual interference: a response planning account. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 51*(4), 735–754. http://dx.doi.org/10.1080/027249898391369.

Jones, L. I., & French, K. E. (2007). Effects of contextual interference on acquisition and retention of three volleyball skills. *Perceptual and Motor Skills, 105*, 883–890.

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

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Koedinger, K. R. (2002). Toward evidence for instructional design principles: examples from Cognitive Tutor Math 6. In D. S. Mewborn, P. Sztajn, D. Y. White, H. G. Wiegel, R. L. Bryant, & K. Nooney (Eds.), (2002). *Proceedings of twenty-fourth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Vol. 1* (pp. 21–49). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review, 19*(3), 239–264. http://dx.doi.org/10.1007/s10648-007-9049-0.

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

Koedinger, K. R., & Corbett, A. (2006). *Cognitive tutors: Technology bringing learning sciences to the classroom*. New York, NY, US: Cambridge University Press.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. The knowledge-learning-instruction (KLI) framework: bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, in press.

Koedinger, K., & Nathan, M. (2004). The real story behind story problems: effects of representations on quantitative reasoning. *Journal of the Learning Sciences, 13*(2), 129–164. http://dx.doi.org/10.1207/s15327809jls1302_1.

Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction, 13*(2), 205–226. http://dx.doi.org/10.1016/S0959-4752(02)00021-X.

Lamon, S. (2005). *Teaching fractions and ratios for understanding: Essential content knowledge and instructional strategies for teachers* (2 ed.). US: Lawrence Erlbaum Assoc Inc.

Lappan, G., Fey, J. T., & Fitzgerald, W. M. (1998). *Bits & pieces: Understanding rational numbers (Connected math series: Number)*. Dale Seymour Publications.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 730–746. http://dx.doi.org/10.1037//0278-7393.9.4.730.

Lee, T. D., & Magill, R. A. (1985). Can forgetting facilitate skill acquisition? In D. Goodman, R. B. Wilberg, & I. M. Franks (Eds.), *Differing perspectives in motor learning, memory, and control* (pp. 3–22) Amsterdam, the Netherlands: Elsevier.

Lee, T. D., & Simon, D. A. (2004). Contextual interference. In A. M. Williams, & N. J. Hodges (Eds.), *Skill acquisition in sport: Research, theory and practice* (pp. 29–44). London, United Kingdom: Routledge.

Lewis, D., & Barron, A. (2009). Animated demonstrations: evidence of improved performance efficiency and the worked example effect. *Human Centered Design, Lecture Notes in Computer Science, 5619*, 247–255. http://dx.doi.org/10.1007/978-3-642-02806-9_29.

Lewalter, D. (2003). Cognitive strategies for learning from static and dynamic visuals. *Learning and Instruction, 13*(2), 177–189. http://dx.doi.org/10.1016/S0959-4752(02)00019-1.

Li, Y. H., & Wright, D. L. (2000). An assessment of the attention demands during random- and blocked-practice schedules. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology, 53*(2), 591–606. http://dx.doi.org/10.1080/027249800390628.

Mack, N. K. (1995). Confounding whole-number and fraction concepts when building on informal knowledge. *Journal for Research in Mathematics Education, 26*(5), 422–441. http://dx.doi.org/10.2307/749431.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(6), 1423–1442. http://dx.doi.org/10.1037//0278-7393.22.6.1423.

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology, 41*(3), 211–253. http://dx.doi.org/10.1006/cogp.2000.0736.

Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: a new model and an experimental curriculum. *Journal for Research in Mathematics Education, 30*(2), 122–147. http://dx.doi.org/10.2307/749607.

Moyer, P., Bolyard, J., & Spikell, M. A. (2002). What are virtual manipulatives? *Teaching Children Mathematics, 8*, 372–377.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

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

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

National Research Council. (2002). *Scientific research in education*. Washington, DC: National Academy Press.

National Science Teachers Association. (2009). *Standards for science teacher preparation*. Revised (2003). National Science Teachers Association, Retrieved 14.11.11, from. http://www.nsta.org/preservice.

Nelson, D. W. (2006). *Effects of practice sequence variations on the transfer of complex cognitive skills practiced in computer-based instruction*. Florida State University.

Ollis, S., Button, C., & Fairweather, M. (2005). The influence of professional expertise and task complexity upon the potency of the contextual interference effect. *Acta Psychologica, 118*(3), 229–244. http://dx.doi.org/10.1016/j.actpsy.2004.08.003.

Paas, F. G. W. C., & van Merrienboer, J. J. G. (1993). The efficiency of instructional conditions: an approach to combine mental-effort and performance measures. *Human Factors, 35*(4), 737–743.

Paas, F., & Van Gog, T. (2006). Optimising worked example instruction: different ways to increase germane cognitive load. *Learning and Instruction, 16*(2), 87–91. http://dx.doi.org/10.1016/j.learninstruc.2006.02.004.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. P. (2007). Enhancing learning and retarding forgetting: choices and consequences. *Psychonometric Bulletin & Review, 14*(2), 187–193. http://dx.doi.org/10.3758/BF03194050.

Rasch, T., & Schnotz, W. (2009). Interactive and non-interactive pictures in multimedia learning environments: effects on learning outcomes and learning efficiency. *Learning and Instruction, 19*(5), 411–422. http://dx.doi.org/10.1016/j.learninstruc.2009.02.008.

Rau, M. A., Aleven, V., & Rummel, N. (2009). Intelligent tutoring systems with multiple representations and self-explanation prompts support learning of fractions. In V. Dimitrova, R. Mizoguchi, & B. du Boulay (Eds.), *14th International conference on artificial intelligence in education* (pp. 441–448). Amsterdam: IOS Press.

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

Reimer, K., & Moyer, P. S. (2005). Third-graders learn about fractions using virtual manipulatives: a classroom study. *Journal of Computers in Mathematics and Science Teaching, 24*(1), 5–25.

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. T. (2007). Cognitive Tutor: applied research in mathematics education. *Psychonomic Bulletin & Review, 14*(2), 249–255. http://dx.doi.org/10.3758/BF03194060.

Ritter, S., Kulikowich, J., Lei, P.-W., McGuire, C. L., & Morgan, P. (2007). What evidence matters? A randomized field trial of cognitive tutor algebra I. *Supporting Learning Flow Through Integrative Technologies, 126*, 13–20.

Rittle-Johnson, B., & Koedinger, K. R. (2005). Designing better learning environments: knowledge scaffolding supports mathematical problem solving. *Cognition & Instruction, 23*, 313–349. http://dx.doi.org/10.1207/s1532690xci2303_1.

Rogers, Y. (1999). What is different about interactive graphical representations? *Learning and Instruction, 9*, 419–425.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics practice problems boosts learning. *Instructional Science, 35*, 481–498. http://dx.doi.org/10.1007/s11251-007-9015-8.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207–217. http://dx.doi.org/10.1111/j.1467-9280.1992.tb00029.x.

Schneider, W. (1985). Training high-performance skills - fallacies and guidelines. *Human Factors, 27*(3), 285–300.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*(2), 141–156. http://dx.doi.org/10.1016/S0959-4752(02)00017-8.

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction, 13*(2), 227–237. http://dx.doi.org/10.1016/S0959-4752(02)00022-1.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on acquisition, retention and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 179–187. http://dx.doi.org/10.1037/0278-7393.5.2.179.

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology, 27*(4), 907–912. http://dx.doi.org/10.1037//0278-7393.27.4.907.

Suh, J., Moyer, P. S., & Heo, J.-J. (2005). Examining technology uses in the classroom: developing fraction sense using virtual manipulative concept tutorials. *Journal of Interactive Online Learning, 3*(4), 1–21.

Sweller, J. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology; General, 119*(2), 176–192. http://dx.doi.org/10.1037//0278-3445.119.2.176.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*(6), 837–848. http://dx.doi.org/10.1002/acp.1598.

Thußbas, C. (2001). Beeinflussen Cover Stories mathematischer Texaufgaben analogen Transfer in der Mathematik? [Do cover stories of text-based math problems influence analogical transfer in mathematics?] *Zeitschrift für Experimentelle Psychologie, 48*(4), 272–289. http://dx.doi.org/10.1026//0949-3946.48.4.272.

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

Van Gog, T., & Paas, F. (2008). Instructional efficiency: revisiting the original construct in educational research. *Educational Psychologist, 43*, 1–11. http://dx.doi.org/10.1080/00461520701756248.

Van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. G. W. C. (2002). Redirecting learners' attention during training: effects on cognitive

load, transfer test performance and training efficiency. *Learning and Instruction, 12*(1), 11–37. http://dx.doi.org/10.1016/S0959-4752(01)00020-2.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227–265.

VanLehn, K., Koedinger, K., Skogsholm, A., Nwaigwe, A., Hausmann, R. G. M., Weinstein, A., et al. (2007). What's in a step? Toward general, abstract representations of tutoring system log data. In C. Conati, & K. McCoy (Eds.), *Proceedings of user modeling.*

Wrisberg, G. A., & Liu, Z. (1991). The effect of contextual variety on the practice, retention, and transfer of an applied motor skill. *Research Quarterly for Exercise and Sport, 62,* 406–412.

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychological Bulletin & Review, 9,* 185–211. http://dx.doi.org/10.3758/BF03196276.

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