Why Interactive Learning Environments Can Have It All: Resolving Design Conflicts Between Competing Goals

Martina A. Rau
Carnegie Mellon Univ.
5000 Forbes Ave
Pittsburgh, PA 15213
marau@cs.cmu.edu

Vincent Aleven
Carnegie Mellon Univ.5000 Forbes Ave
Pittsburgh, PA 15213
aleven@cs.cmu.edu

Nikol Rummel
Ruhr- Univ. Bochum
44780 Bochum, Germany
nikol.rummel@rub.de

Stacie Rohrbach
Carnegie Mellon Univ.
5000 Forbes Ave
Pittsburgh, PA 15213
stacie@cmu.edu

ABSTRACT

Designing interactive learning environments (ILEs; e.g., intelligent tutoring systems, educational games, etc.) is a challenging interdisciplinary process that needs to satisfy multiple stakeholders. ILEs need to function in real educational settings (e.g., schools) in which a number of goals interact. Several instructional design methodologies exist to help developers address these goals. However, they often lead to conflicting recommendations. Due to the lack of an established methodology to resolve such conflicts, developers of ILEs have to rely on ad-hoc solutions. We present a principled methodology to resolve such conflicts. We build on a well-established design process for creating Cognitive Tutors, a highly effective type of ILE. We extend this process by integrating methods from multiple disciplines to resolve design conflicts. We illustrate our methodology's effectiveness by describing the iterative development of the Fractions Tutor, which has proven to be effective in classroom studies with 3,000 4th-6th graders.

Author Keywords

Design conflicts; instructional design; interactive learning environments; Cognitive Tutors.

ACM Classification Keywords

K.3.1 [Computers and Education]: Computer Uses in Education – Computer-assisted instruction

H.5.2 [Information Interfaces and Presentation]: User Interfaces - User-centered design

General Terms

Human Factors; Design; Experimentation.

INTRODUCTION

To design interactive learning environments (ILEs), instructional designers identify stakeholder goals and address them within the constraints of the given educational context. Often, stakeholder goals can readily be identified as they are in part dictated by state requirements (e.g., performance on standardized tests) and the context (e.g., the necessity for classroom management). In addressing these goals, developers can draw on many frameworks which guide the design of ILEs to identify and address stakeholders' goals while matching resource limitations [1-7].

But in reality, it is inevitable that different stakeholders have different goals, that there are significant resource limitations, and that design goals (even if they were agreed upon by all stakeholders) need to be traded off against each other. Unfortunately, existing design frameworks for ILEs do not address how to resolve such design conflicts. This may in part be due to the fact that different types of design frameworks are relevant to the development of ILEs: some frameworks focus on user-centered design [1-2, 4], others incorporate learning science [3] and educational psychology research [5-7]. However, these different types of frameworks rarely reference one another. For this reason, developers often have to rely on ad-hoc methods to resolve conflicts that arise in the interdisciplinary field of ILEs. For instance, a math teacher who wants to help students learn deeply may provide complex real-world problems [3]. Yet, [7] suggest to practice part-tasks: discrete tasks that are necessary for the completion of complex problems (e.g., practicing math facts). At the same time, students find complex problems interesting, but the teacher might worry that their learning is jeopardized because the problems do not provide just-in-time feedback [7].

It crucial that we pay attention to the conflicting goals that inevitably occur in complex educational settings. If we fail to address stakeholders' competing goals, our ILE will never be as successful as it can be: students may dislike the ILE because it is either boring or too challenging, or teachers – who might well believe it will help their students learn deeply – fail to use the ILE within the constraints of their day-to-day job which requires them to prepare students for standardized tests and manage a class of students. However, if we succeed in integrating stakeholders' needs within the constraints of their contexts into the design of our ILE, the dissemination of our product will hugely benefit.

Nevertheless, to integrate competing goals and constraints, developers of ILEs typically have to rely on ad-hoc approaches to resolve design conflicts. What is needed is a principled methodology that developers can apply to resolve such conflicts. The goal of the present paper is to describe a new approach to resolving conflict that arise between multiple goals and constraints in educational settings.

CHI 2013, April 27–May 2, 2013, Paris, France.

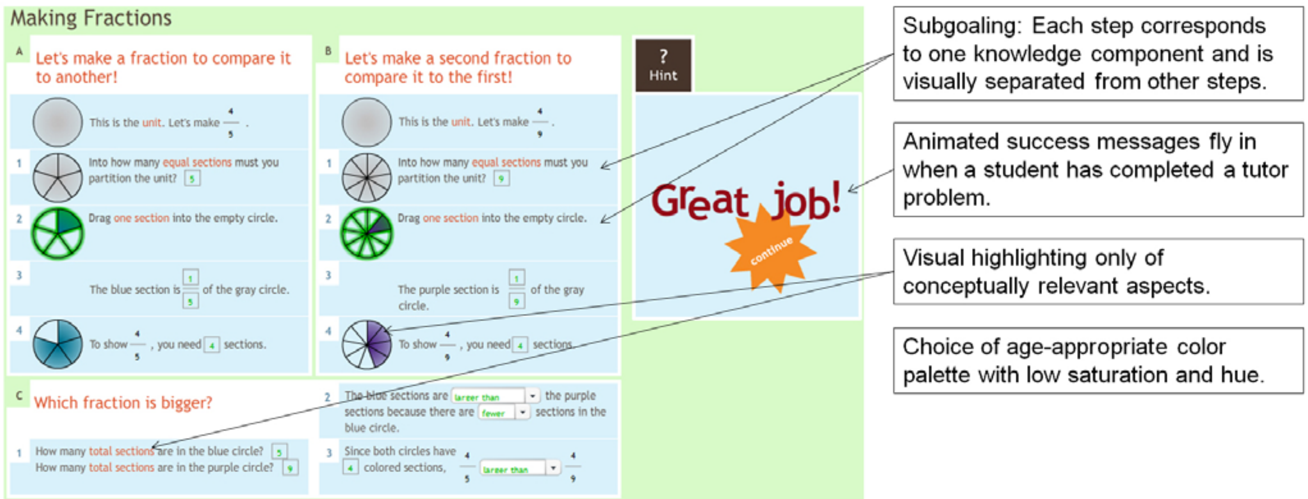

Figure 1. Example problem in the Fractions Tutor that employs the subgoaling strategy.

We illustrate our multi-method approach by describing our research within a particularly successful ILE: Cognitive Tutors. Cognitive Tutors are grounded in cognitive theory and artificial intelligence. They pose rich problem-solving tasks to students and provide individualized support at any point during the problem-solving process. At the heart of the Cognitive Tutors lies a cognitive model of students' problem-solving steps. The model is used to provide individualized support for students during the learning process [8-10]. Cognitive Tutors have been shown to lead to significant learning gains in a variety of studies [8-12]. They are currently being used in close to 3,000 U.S. schools. Cognitive Tutors are particularly suitable for describing our approach to resolving design conflicts as their development follows a well-described design process that integrates design recommendations originating from a number of fields, including HCI, learning science, and education [13]. We extend this process by providing a new approach for resolving conflicting design recommendations and constraints. In particular, our methodology combines focus groups and affinity diagramming to develop a goal hierarchy, parametric experiments, and cross-iteration studies. The novelty of our approach lies in a principled combination of methods that originate in a variety of disciplines, including from HCI, learning sciences, education, and ILE design.

We illustrate the success of our methodology using the Fractions Tutor as example (see Figure 1) – a successful Cognitive Tutor that has gone through several iterations of classroom experiments and lab studies with over 3,000 students. The Fractions Tutor has been shown to significantly enhance students' learning. In a recent classroom study with a total of 599 4th- and 5th-graders [14], after 10 hours of instruction with the Fractions Tutor, students improved significantly with a medium effect size of d = .40 at the posttest ($p < .01$). When we administered a delayed posttest a week later, we found that students retained these learning gains with an effect size of d = .60 ($p < .01$).

The success of the Fractions Tutor, like that of other Cognitive Tutors [15-17], has been shaped by incorporating stakeholder goals into the design process. In doing so, we employed a principle-based approach which we believe is not unique to the domain of fractions or Cognitive Tutors in particular, but that can inform design decisions made during the development of a wide range of ILEs.

COGNITIVE TUTOR DESIGN PROCESS

Before we present our novel approach to resolving conflicts in designing ILEs, we review the development process for Cognitive Tutors [5, 8-9]. This process comprises a set of iterative, non-linear stages.

Stage 1: Stakeholder and Problem Identification

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

The development of the Fractions Tutor was motivated by the fact that students struggle with fractions as early as elementary school [18-19]. Fractions are considered an important educational goal in and of itself, and an important prerequisite for later algebra learning [18]. Interviews with teachers confirmed the need for an effective ILE that can help students overcome their difficulties with fractions.

Stage 2: Identifying Assessment and Practice Problems

Based on the educational problem, designers should identify a set of assessment tasks (i.e., tasks learners should be able to solve after having worked with the ILE). Assessment tasks should guide the selection of practice problems (i.e., problems students should solve as part of the ILE). A search of the education literature will yield a set of domain-specific target problems both for assessment and for practice. In addition, developers and teachers should brainstorm about novel problems for assessment and practice.

The outcome of stage 2 is a set of domain-specific assessment tasks and practice problems.

Stage 3: Cognitive Task Analysis

The goal of stage 3 is to understand student learning and student thinking in the domain. In doing so, we identify the knowledge and strategies the ILE should cover, using cognitive task analysis techniques [21]. Cognitive task analysis seeks to identify the knowledge components (i.e., units of knowledge) that students need to acquire to perform well on assessment and practice problems. Cognitive task analysis is based on think-alouds and observations of student learners (novices) or proficient student (experts), or based on a theory of what knowledge learners need to acquire. Think-alouds and observations are often combined with difficulty factors assessment [22] – a method to identify features of tasks that reliably change the difficulty of the task.

Several iterations between stages 2 and 3 are recommended. After each iteration, developers should update the collection of assessment and practice problems based on insights gained from cognitive task analysis and difficulty factors assessments. They should review the problems in focus groups or in interviews with teachers, while also discussing problem sequences within the ILE. Observations of teacher-student tutoring can help guide the instructional design, by providing insights into successful instructional strategies.

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

Stage 4: Cognitive Modeling and Tutor Development

Stage 4 aims at developing the Cognitive Tutor. As part of this stage, developers will create the Cognitive Tutor interface, a cognitive model of student problem solving that will serve as a basis for individualized support, and a curriculum that contains a collection of problem types and that span across a variety of topics within the given domain.

Typically, stage 4 includes several cycles of rapid, low-fidelity prototyping and high-fidelity prototyping. These rounds of testing are usually conducted in the laboratory with a small number of students from the target population. Between each round of testing, the materials are updated based on the findings and issues identified. To develop the Fractions Tutor, we used the Cognitive Tutor Authoring Tools (CTAT, [23]), which allows for rapid prototyping and fast implementation of iterative design changes. We recommend including teachers in the development process. Involving teachers will not only improve the quality of the ILE – it will also benefit the dissemination of the final product and is likely to reveal further stakeholder goals.

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

Stage 5: Pilot studies and parametric studies

The goal of stage 5 is to formally evaluate and iteratively improve the ILE. A set of methods are available during this phase. First, pilot testing in the laboratory is useful to get in-depth insights with students solving practice problems while thinking aloud, which can help identify gaps in their knowledge that the ILE does not yet address. Second, testing the ILE in classrooms is indispensible. We can gather a variety of data from these classroom studies. Students' learning gains should be assessed based on pretests and posttests that integrate the target problems identified during earlier stages, including both standardized test items and transfer items that assess students' ability to apply their knowledge to novel task types. Informal observational data of students' interactions with the ILE in classrooms, interviews and focus groups with teachers as well as surveys with students and with teachers will yield valuable insights into usability issues and reveal crucial aspects of the stakeholders' goals. Log data gathered while students use the ILE provides a useful basis for identifying issues in usability and difficulty level of particular steps within the ILE.

Since fractions instruction typically uses a variety of graphical representations (e.g., circles, rectangles, and number lines) [18-20**,** 24], the Fractions Tutor includes multiple interactive graphical representations. To inform our design choices in integrating these graphical representations, we conducted a series of parametric studies in classrooms. As a consequence of these studies, the Fractions Tutor uses graphical representations in the following manner:

- The Fractions Tutor switches frequently between different graphical representations, as opposed to providing several practice problems with one graphical representation before moving on to another graphical representation.
- The Fractions Tutor provides support for relating graphical representations (e.g., circles) to symbolic representations (e.g., ½) in the form of menu-based self-explanation prompts.
- The Fractions Tutor provides support for relating different graphical representations (e.g., circles and number lines) that requires students to become active in making these relations, rather than having the ILE provide these connections automatically.

The outcome of stage 5 is a set of updated stakeholder goals, as well as an updated and iteratively improved ILE that is ready for classroom dissemination.

Stage 6: Classroom Use and Evaluation

After several iterations with the ILE, it is time to evaluate the system. Randomized field trials are the method of choice during this phase. A large number of classrooms should be randomly assigned to using the ILE or to work with another commonly used type of ILE, or a curriculum without an associated ILE. The success of the ILE should not only be evaluated based on students' performance on pretests and posttests. Observations in randomly selected classrooms, interviews with randomly selected teachers and student or teacher surveys will identify problem-solving behaviors and learning processes. Further, the analysis of student log data during problem solving can help detect issues with specific problem-solving steps, for example by identifying steps on which students make many errors.

IDENTIFYING STAKEHOLDER GOALS AND INSTRUCTIONAL DESIGN PRINCIPLES

We now describe our novel contributions to the design processes just described. We formulate a hierarchy of stakeholder goals. Then, we identify the instructional design recommendations to address these goals and identify conflicts between instructional design recommendations. Finally, we present three approaches to resolving these conflicts.

Forming a Goal Hierarchy

Across the stages and iterations of ILE development, one must not lose sight of stakeholder goals. We recommend conducting focus groups and interviews with teachers and students as part of each stage. Further, we suggest creating a goal hierarchy to identify and resolve goal conflicts.

To develop a goal hierarchy, we use affinity diagrams, a common HCI technique [25]: we write each goal on a sticky note and then work bottom-up to organize them into a hierarchy. Once all notes are collected in groups, we name the group. We then identify a set of instructional design recommendations that can help us achieve each goal.

Identifying Goals and Design Recommendations

Table 1 provides an overview of the goals and instructional design recommendations we identified for the Fractions Tutor. We identified the overarching goal to support students' robust learning of fractions (G1). Based on these interviews and based on the review of of educational standards [18-20], we identified teachers' goals to promote students' learning of robust knowledge about fractions which can transfer to new problem types and that lasts over time (G1). Based on the education standards and mathematics literature, we formulated domain-specific goals to promote conceptual understanding of fractions as parts of a whole, as proportions, and as measurements. Both the mathematics education literature and the learning sciences literatures suggest that instruction should employ graphical representations to illustrate these different interpretations of fractions (id1) [18-20,24]. Furthermore, education standards describe the need for students to learn multiple strategies to solve fractions problems (id2) [18-20]. For instance, students should be able to compare fractions with like numerators or like denominators by reasoning about the relative size of fractions, and using benchmarks and equivalent fractions. To enhance learning of a variety of strategies, the education literature suggests practicing different strategies to solve the same problem [26] and to discuss why they lead to the same [27]. Further, students should be exposed to a variety of problem types [28], which should enhance their ability to transfer their knowledge to novel problem formats (id3). To further enhance robust learning, the learning sciences result literature recommends using complex realistic problems with cover stories (id4) [29], to introduce encourage abstraction (id5) [28], to illustrate the structural components

Goals	Instructional Design Principles	Conflicts
G1: Robust domain knowledge that lasts and transfers to new problems types	G1: use variety of graphics (id1), use a variety of problem formats (id2), practice variety of strategies (id3), use cover stories (id4), introduce abstract terms (id5), subgoaling, (id6), use lean design that highlights only conceptually relevant aspects (id7), use complex holistic problems (id8)	C1 (within G1): provide complex, holistic problems (id8) vs. use subgoaling (id6)
G2: Performance on standardized tests	G2: solve problems that are structurally similar to the target problems (id9), provide opportunities for extended practice to master one strategy (id10)	C2a (with G1): practice with a variety of problem formats (id2) vs. practice with problem formats that are structurally similar to the target problems (id9) C2b (with G1): practice a variety of strategies (id3) vs. rote practice of one strategy that is most likely to always lead to the correct answer (id10)
G3: Ease of classroom management	G3: engaging, usable system (id12), gather information about students' performance (id13), self-paced learning and individualization (id15)	C3 (with G1): provide complex real-world problems with cover stories (id4) vs. provide an engaging, easy-to-use system (id12)
G4: Tasks that invoke self-efficacy	G4: employ concrete language that students can understand intuitively (id11)	C4: use abstract language that applies to a variety of situations (id6) vs. use of intuitive language that uses concrete examples (id11)
G5: Fun, interest, entertainment	G5: game-like colorful and flashy elements (id14)	C5: include game-like, colorful elements whose main purpose is to visually appeal young students (id14) vs. lean designs that use colorful highlighting to emphasize conceptually relevant aspects (id7)

Table 1. Goals and Instructional Design Principles.

of a problem-solving procedure using subgoaling (id6) [30]. Subgoaling is a procedure that aims at communicating the goal structure of a problem by breaking it into clear substeps, thereby "making thinking visible." Instructional design principles suggest to use color only sparingly, and to highlight only conceptually relevant aspects of the problem (id7) [6].

Our focus groups, surveys, and interviews with teachers revealed a further important goal: to help students perform well on standardized tests (G2). To address this goal, one might suggest to provide opportunities for practice on problems that are structurally similar to the target assessment problems (id9) [31], and to master one strategy that will lead to the correct answer (id10) - to the extent one can predict the kinds of items that appear on such tests.

Our classroom observations demonstrated teachers' needs for classroom management while using the Fractions Tutor (G3), including the ability to focus on students who struggle with the content, monitoring students' progress, and a quiet classroom of students who concentrate on their work. For instance, when asked what they like about using ILEs, teachers reported: "I like using it because it is so interactive for the students. They stay very involved," or "The programs that I use with my students are interactive, colorful, and can hold their attention." To address teachers' goal for classroom management, we used focus groups with teachers to identify possible obstacles that our system created within the classroom. We discovered that any aspect that makes the ILE difficult to use for students results in teachers helping students out with usability issues rather than helping with the content. An ILE that is easy to use and that includes easy math problems (see G4) would thus help achieve this goal (id12). Furthermore, teachers expressed their interest in a system that would gather real-time information about individual students' performance on the practice problems, so that they could more easily identify struggling students and help them out (id13).

Furthermore, teachers' concerns about the difficulty of practice problems, in particular of complex real-world problems, revealed their goal of using a low level of difficulty (G4). Think-alouds and interviews with students further demonstrated that students value easy tasks that invoke feelings of self-efficacy; i.e., tasks that make them feel like they can do math. To make problems easier for students, instructional materials should employ concrete language that students can understand intuitively (id11).

Finally, surveys with students demonstrated their goal to have fun and to be entertained (G5). This need might best be achieved focusing on age-appropriate design elements resembling games with colorful and flashy elements (id14). Also, complex real-world problems might address students' need for interesting practice problems (see id4).

Forming a Hierarchy of Stakeholder Goals

Next, we create a hierarchy of the goal categories we identified. In doing so, focus groups with the stakeholders can help inform the ranking of goals. In our case, an important question regarded the relative importance of the goal to promote robust learning (G1) and the goal to help students perform well on standardized tests (G2). As mentioned, these goals do not always result in competing design recommendations: The purpose of standardized tests is to assess students' robust learning. Practicing with only one format of questions may be detrimental to students' performance on the test if the test contains questions other than the ones that were anticipated. Since our interviews and focus groups with teachers demonstrated that they cared deeply about students' deep learning, we felt that the stakeholders' goals were best represented by ranking the goal to promote students' robust learning (G1) higher than the goal to do well on standardized tests (G2).

Goal	Effects	Impact
G3: Classroom management	Information about which students are struggling allows teachers to help them out, and increases these students' learning *(+ learn)*	1 + 1 = *2*
	A quiet classroom allows students to concentrate better and will help them learn *(+ learn)*	
G4: Self-efficacy	Easy problems increase students' enjoyment *(+ G5: fun)*	0.5(-1 + 1 + 1 – 1) + 1 – 1 + 0.5(1+1) = *0.5*
	Easy, but somewhat challenging problems enhance students' learning *(+ learn)*	
	Too easy problems do not require deep processing of the content and do not help students learn *(- learn)*	
	Students have less trouble with easy problems, which positively affect classroom management *(+ G3: management)*	
G5: Fun	Distracting elements are known to impede learning *(- learn)*	-1 + 1 + 1 – 1 = *0*
	Fun software is more motivating, which might enhance learning *(+ learn*	
	Fun software may help classroom management if it motivates students to work on the task *(+ learn)*	
	Fun software might impede classroom management if it decreases students' concentration *(- learn)*	

Table 2. Impact Factors of Stakeholder Goals.

Understanding the context of our ILE can also crucially inform our goal hierarchy. In schools, teachers and superintendents are the ones who decide to acquire and use educational software. Performance on standardized tests is one highly important metric on which teachers are evaluated. In other words, the ability to promote performance on standardized tests is a prerequisite for the dissemination of ILEs. We therefore consider performance on standardized tests (G2) as highly important.

For the goals on which we cannot find consensus in focus groups (i.e., G3-5), we again use affinity diagrams to identify classes of goals based on the effect they have on students' learning and on the dissemination of the ILE. In doing so, a brainstorming session with experts (who have good knowledge of the relevant literatures) about the effects of common interventions to meet the goals can help create the goal hierarchy. We then regroup the generated items to create a diagram for the effects. The resulting categories are provided in italics in the list below. We then computed the impact of each goal. To do so, we assigned a value of 1 to each positive effect (e.g., the intervention will promote learning) and a value of -1 to each negative effect (e.g., the intervention will harm students learning). To account for effects of one goal on another related goal (e.g., if one of the effects of G4, providing easy tasks is to enhance G5, having fun), included the effects of the second goal (e.g., + fun = -1 + 1 + 1 - 1) while discounting them by 0.5 (since they constitute an indirect effect). Table 2 provides an overview of the effects we identified for each goal and of their impact factor. We note that the calculus in Table 2 can be used as a guideline: more crucial than the actual impact factors is the consideration of the specific effects of achieving goals and the interactions among multiple goals.

Based on these impact factors we update the goal hierarchy: after robust learning (G1) and performance on standardized tests (G2), classroom management (G3) is the next-most important goal, followed by easy tasks (G4), and fun (G5).

Conflicts between design recommendations

We will now turn to mapping out the conflicts that arise from competing goals and from the resulting instructional design recommendations described in the previous section. To identify these conflicts, we recommend to conduct focus groups with learning sciences experts who have in-depth knowledge of the empirical research on the various design recommendations. Table 1 summarizes the conflicts we identified based on this method for the Fractions Tutor.

A number of conflicts arise between the various instructional design recommendations that address the goal to promote the learning of robust domain knowledge (G1), and with teachers' requirements to enhance students' performance on standardized assessments (G2). The need to perform well on standardized assessments does not always dictate approaches that compete with deep learning – but when such conflicts do arise, they are not trivial, and being aware of them is crucial. One such *conflict C2a* exists between the goal to promote the learning of robust domain knowledge (G1) through practice with a variety of problem formats (id2) and the goal to enhance students' performance on standardized tests (G2) through practice with problem formats that are structurally similar to the target problems (i.e., standardized test items, to the extent they are known, id9). Another *conflict C2b* exists between the recommendation promote robust learning (G1) to practice a variety of different strategies (id3), and to enhance students' performance on standardized tests (G2) through practice of one strategy that is most likely to always lead to the correct answer (id10).

Conflicts also arise between design recommendations that address the same goal. One such *conflict C1* exists within the goal to promote robust learning (G1) by providing holistic and complex problems (id8) or by using subgoaling to break the problem up into small steps (id4).

Further conflicts can arise from constraints within schools and students' abilities. For example, before the background of students' poor reading ability, *conflict C3* occurs between the goal to promote robust learning (G1) by providing complex real-world problems with cover stories (id4) and teachers' needs to facilitate classroom management (G3) by providing an easy-to-use system (id12): In our own classroom studies, we found that teachers were busy helping students understand problem statements which were lengthy due to the realistic cover stories the tutoring system provided.

Finally, conflicts arise between design recommendations on how to promote robust learning (G1) and students' emotional and cognitive needs. One such *conflict C4* results from promoting robust learning (G1) through the use of abstract language that applies to a variety of situations (id6) and students' need for easy problems (G4) through the use of intuitive language that uses concrete examples (id11). Another *conflict C5* exists between students' preference for flashy designs recommend the inclusion of game-like elements whose main purpose is to visually appeal to young students (id14), and the use of lean designs so as to not distract the user from the learning task, and that use colorful highlighting only sparingly to emphasize conceptually relevant aspects (id7).

RESOLVING CONFLICTS

In the following, we provide a principled process to address the conflicts we identified. In doing so, we describe three approaches: (1) resolve conflicts based on the goal hierarchy where possible, (2) conduct parametric experiments, and (3) conduct cross-iteration studies. Although we present these three approaches as a sequence, they complement each other. Crucial to the success of the conflict resolution is the careful evaluation of design solutions based on lab studies and classroom evaluations. In this sense, conflict resolution is most likely to occur as part of stage 5 in the Cognitive Tutor design process described above.

Resolving Conflicts Based on the Goal Hierarchy

First, we use the goal hierarchy to resolve some of the conflicts we previously identified. Conflict C2a exists between the goal to promote robust learning (G1) and the goal to promote performance on standardized tests (G2) (i.e., practice with a variety of problem formats (id2) versus practice with problem formats that are structurally similar to the target problems (id9)). Both goals are important, but we prioritize the goal to promote robust learning (G1). We thus recommend to practice with a variety of formats (id2), rather than to practice with only the problem format that is structurally equivalent to the test format (id9). However, practice items that structurally correspond to the test format should also be included in the variety of practice items.

A similar conflict C2b exists between the goal to promote robust learning (G1) and to promote performance on standardized tests (G2) (practice a variety of different strategies (id3) versus practice of one strategy that is most likely to always lead to the correct answer (id10)). Here again, we recommend giving larger weight to robust learning (G1) and practice of a variety of strategies (id3) than to practicing only one strategy (id10). However, even while practicing a variety of strategies, we can prioritize one strategy, especially when using mastery learning within Cognitive Tutors. Mastery learning employs a Bayesian decision process to decide whether students should continue solving problems to practice the target problem-solving strategy, based on the probability that the student has learned that knowledge [13]. This process can be used to ensure that all students master at least one strategy while also being exposed to multiple strategies.

Another conflict C5 is between robust learning (G1) and students' goal to have fun (G5) (i.e., inclusion of game-like, colorful elements whose main purpose is to visually appeal young students (id14) versus lean designs that use colorful highlighting to emphasize conceptually relevant aspects (id7)). Given that our goal hierarchy places the highest priority on supporting robust learning (G1), whereas the whereas the goal to have fun (G6) has lowest priority, it is clear that we should prioritize on employing color-based highlighting only conceptually relevant aspects. However, we can do so in a way that is visually appealing to students of our target age group. Further, we can integrate flashy and exciting elements where (or when) they do not distract, for instance, at the end of a practice problem.

Figure 1 illustrates several key aspects of the solution we chose for the Fractions Tutor. First, our choice in color reflects the finding that students in grades 4 and 5 have a preference for less intense colors with lower saturation and hue, compared to younger students [32-33]. We also made sure the colors we selected are gender neutral [33]. Second, in the service of using color to emphasize only conceptually relevant aspects, we use orange to highlight key words in each problem step. Finally, Figures 1 and 2 show a success message that we display at the end of a problem. The message contains a short movie clip that flies in.

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

Parametric Experiments to Resolve Conflicts

Conflicts that cannot be resolved based on the goal hierarchy require more careful inspection. We recommend conducting parametric experiments using multiple metrics to address important remaining conflicts. These studies can be carried out as part of stage 5 in the design process described earlier. We addressed conflict C1 between holistic problems and subgoaling as part of a parametric experiment.

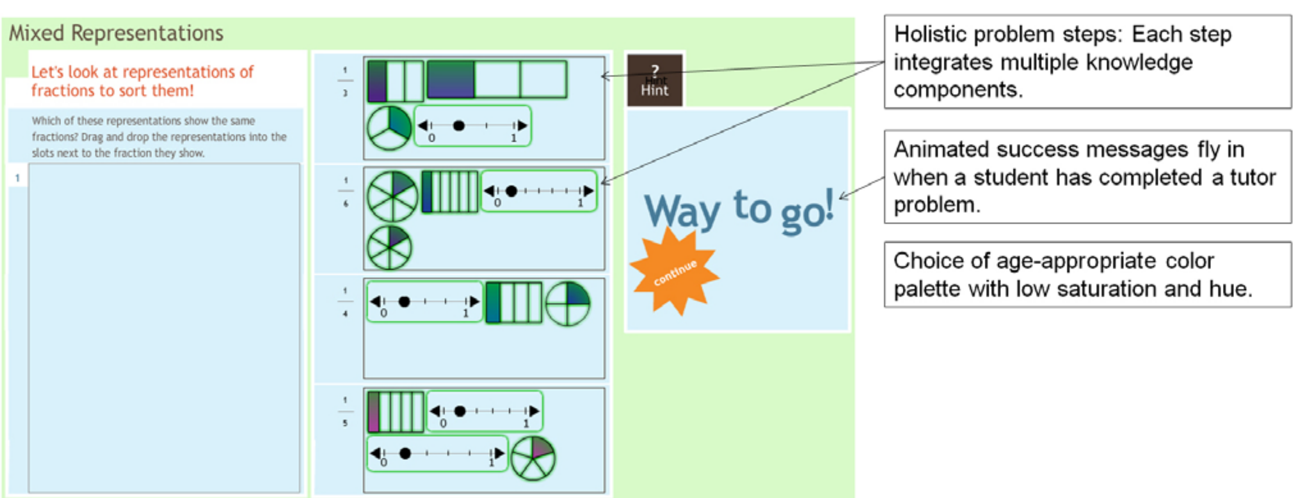

Figure 2. Example Problem in the Fractions Tutor that Employs a Holistic Approach.

As mentioned, the subgoaling strategy (illustrated in Figure 1) breaks up problems into their substeps, in order to communicate the problem's goal structure [29]. However, our surveys show that students tend to dislike multi-step problems. In a survey following an early classroom study in which 311 students worked with the Fractions Tutor, a student commented, for instance: "A suggestions i would make is stop the repeating and give more fun stuff because i heard from people even me not to be mean but most of it ws boring sorry." Another student said: "in my opinion that there were too many questions in one problem!!" Having many steps within a problem seems to overwhelm students. For example, a student reported: "I think there was too many questions." To address this issue, we decided to investigate whether we can enhance students' learning by decreasing the grain-size at which we support their problem solving. In an experimental study that we conducted with 599 students, we investigated the impact of the proportion of tutor problems with subgoaling to those with holistic problems on students' learning and on their enjoyment of the Fractions Tutor. Specifically, we compared versions of the Fractions Tutor in which 100%, 75%, or 50% of practice problems employed the subgoaling strategy (AllSubgoal vs. 75Subgoal vs. 50Subgoal).

Students first took a pretest, then worked with the Fractions Tutor for a total of 10 hours, then took an immediate posttest. One week after the immediate posttest, students took a delayed posttest. The tests comprised test items that assessed students' learning of the content covered in the Fractions Tutor, and their ability to transfer that knowledge to new task types. All tests were equivalent (i.e., they contained the same type of test items, but not identical test items). Students who completed all tests, and who completed their work on the tutoring system were included in the analysis, yielding a total of $N = 428$. The number of students who were excluded from the analysis did not differ between conditions, χ^2 (6, N = 169) = 4.34, $p > .10$. A repeated measures ANCOVA with pretest as covariate, immediate posttest, and delayed posttest as dependent mea sures and condition as independent factor showed a marginally statistically significant main effect for condition, $F(2, 424) = 2.74$, $p = .06$. Figure 3 depicts students' scores on the immediate posttest by condition. Post-hoc comparisons showed that students who worked with the AllSubgoal or the 75Subgoal versions of the Fractions Tutor significantly outperformed students who worked with the 50Subgoal version (ps < .05). There were no statistically significant differences between the AllSubgoal version and the 75Subgoal version ($p = .13$). Results from our survey showed that while 53% of the students reported that they liked the tutor problems with subgoaling, 77% of the students reported that they liked the tutor problems without subgoaling.

Taken together, results from the parametric experiment demonstrate that by including some holistic problems, we could improve students' enjoyment while working with the Fractions Tutor without harming their learning gains. Holistic problems may be more engaging as they resemble real-life tasks which are complex in nature [7]. Further, we identified what proportion of problems can use a holistic approach without harming students' learning: 50% of holistic problems impedes students' learning, as Figure 3 illustrates. However, 25% of problems can take a holistic approach, without hampering learning.

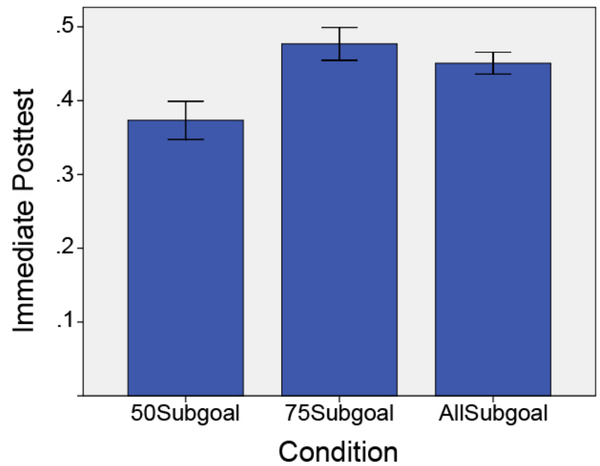

Figure 3. Students' posttest scores by condition.

Cross-Iteration Studies to Resolve Conflicts

Unfortunately, it is not possible to conduct a controlled experiment for every design decision. In this case, we recommend that developers conduct cross-iteration studies. We addressed conflict C3 between the goal to promote robust learning (G1) by providing complex real-world problems with cover stories (id4) and the goal to facilitate classroom management (G3) by providing an easy-to-use system (id12) based on the effects of the design decision across several iterations of the Fractions Tutor.

Initially, we resolved conflict C3 based on our goal hierarchy, which prioritizes robust learning. However, when employing a version of the Fractions Tutor that included cover stories in classrooms, we faced challenging issues. Students complained about having to read a lot, and teachers expressed their concern about being able to use the ILE in their classrooms without extra help. Several teachers suggested including an audio function, so that students could listen to the problem statement via headphones. However, since many schools lack the necessary equipment (i.e., headphones), we discarded that idea. Instead, we decided to exclude cover stories from the Fractions Tutor. However, in a subsequent classroom study, our classroom observations demonstrated that students had trouble making sense of the rather abstract problems in the tutor. An anonymous survey with 331 students revealed that students thought the problems were too hard and that they were not fun. One student commented, for instance: "I don't like how the problem didn't give clear, vivid questions. It confused the way I was taught." Several students commented on the

ILE being boring, for instance: “it was good but it got boring at times.”

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

These cross-iteration changes to the Fractions Tutor illustrate that in cases where design choice based on the goal hierarchy proves to be impractical, several iterations may be necessary to find a balance between the disadvantage of the desired design choice, and alternative solutions. By carefully monitoring the effect of each design choice, we believe that the combination of cover stories in introductory problems and less reading-intensive, abstract problems is an effective and practical solution for the young population the Fractions Tutor is designed for.

CONCLUSION

This paper presents a principled, multi-method approach to resolving the conflicts that inevitably arise between competing design goals in complex educational settings. We provide empirical evidence that our approach lead to the development of a successful ILE which not only significantly improves students’ learning, but which both students and teachers find easy and enjoyable to use.

Although at times, design decisions are situational, highly contextualized and occur under the pressure of deadlines and therefore are bound to be (to some extent) arbitrary, our approach addresses the common scenario in which developers of ILEs need to rely on ad-hoc methods to resolve conflicts between conflicting goals of multiple stakeholders. We combine focus groups and affinity diagramming to develop a goal hierarchy, parametric experiments, and cross-iteration studies, thereby extending existing instructional design processes by integrating methods from multiple disciplines. Specifically, we use a goal hierarchy to resolve conflicts, in combination with parametric experiments and cross-iteration studies.

Although we developed and evaluated our approach within the context of a Cognitive Tutor, a specific type of ILE that is widely used across 3,000 schools in the United States, we are confident that our approach will generalize to other types of ILEs. For instance, MIT’s edX system, an open-source learning technology, which makes course materials at the college level accessible online, faces unique design challenges due to the learners’ contexts and goals: they may be students from around the world using the system for exam preparation, or teachers who access the system in order to fulfill their continued education requirement. Conflicts might exist between the users’ goal to relate the learning content to specific contexts, such as for an engineering project (if the user is a college student majoring in engineering), or for a high-school classroom (if the user is a teacher). Addressing these goals is difficult because tailoring the content to these different interest groups would result in having highly specific content that is not at the same time relevant to all interest groups. Yet, MIT has an interest in the edX system being widely used across different groups of users. In applying our approach to create a goal hierarchy for different types of users, in conducting parametric experiments and cross-iteration studies, trade-offs such as the one just described can be explicitly identified and addressed.

The scenario with edX illustrates that the approach we describe in this paper might serve as a framework to stimulate future research on ILE development, not only to improve specific ILEs, but also to evaluate and further extend the presented approach. Only with a well-researched and principled approach to incorporating multiple (and, as described, often conflicting) stakeholders’ goals can we have it all: popular, usable, and effective ILEs.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation, REESE-21851-1-1121307. We thank Ken Koedinger, Mitchell Nathan, Kathy Cramer, Peg Smith, Jay Raspat, Michael Ringenberg, Brian Junker, Howard Seltman, Cassandra Studer, the students, teachers, and principals, the CTAT and the Datashop teams.

REFERENCES

1. Soloway, E., Jackson, S. L., Klein, J., Quintana, C., Reed, J., Spitulnik, J., Stratford, S. J., Studer, S., Eng, J. and Scala, N. *Learning theory in practice: Case studies of learner-centered design*. In *Proc. CHI ‘96*, ACM Press, (1996).
2. Design-based research collective. An emerging paradigm for educational inquiry. *Educational Researcher, 32*, 1 (2003), 5-8.
3. Bereiter, C. and Scardamalia, M. Learning to work creatively with knowledge. In E. de Corte (Ed.), *Powerful learning environment*, (2003).
4. Jackson, S., Krajcik, J. S. and Soloway, E. The design of guided learner-adaptable scaffolding in interactive learning environments. In *Proc. CHI ‘98*, ACM Press, (1998).
5. Koedinger, K. R. *Toward evidence for instructional design principles: Examples from Cognitive Tutor Math 6*. In *Proc. PME/NA XXXII* (2002).

6. Mayer, R. E. The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction*, 13, 2 (2003), 125-139.
7. van Merrienboer, J. J. G., Clark, R. E. and de Croock, M. B. M. Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development*, 50, 2 (2002), 39-64.
8. Koedinger, K. and Corbett, A. *Cognitive Tutors: Technology bringing learning science to the classroom.* Cambridge University Press, 2006.
9. Corbett, A. and Anderson, J. Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In *Proc CHI 2001,* ACM Press, (2001).
10. Corbett, A. and Trask, H. Instructional interventions in computer-based tutoring: Differential impact on learning time and accuracy. In *Proc CHI 2000,* ACM Press, (2000).
11. Corbett, A. T., Koedinger, K. and Hadley, W. S. *Cognitive tutors: From the research classroom to all classrooms.* Lawrence Erlbaum Associates, 2001.
12. VanLehn, K. The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educational Psychologist*, *46*, 4 (2011), 197-221.
13. Corbett, A., Koedinger, K. R. and Anderson, J. R. *Intelligent tutoring systems*. Elsevier Science Publishers, 1997.
14. Rau, M. A., Aleven, V., Rummel, N., Rohrbach, S. Sense Making Alone Doesn’t Do It: Fluency Matters Too! ITS Support for Robust Learning with Multiple Representations. In *Proc ITS 2012,* Springer, (2012).
15. Ogan, A., Aleven, V. and Jones, C. Pause, predict, and ponder: Use of narrative videos to improve cultural discussion and learning. In *Proc CHI 2008,* ACM Press, (2008).
16. Corbett, A., Kauffman, L., MacLaren, B., Wagner, A. and Jones, E. A cognitive tutor for genetics problem solving: Learning gains and student modeling. *Journal of Educational Computing Research*, *42*, 2 (2010), 219-239.
17. Ritter, S., Anderson, J. R., Koedinger, K. R. and Corbett, A. T. Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, *14*, 2 (2007), 249-255.
18. Siegler, R. S., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., Thompson, L. and Wray, J. *Developing effective fractions instruction: A practice guide*. U.S. Department of Education, 2010.
19. NMAP. *Foundations for Success: Report of the National Mathematics Advisory Board Panel*. U.S. Government Printing Office, 2008.
20. NCTM *Principles and Standards for School Mathematics*. National Council of Teachers of Mathematics, 2000.
21. Schraagen, J. M., Chipman, S. F. and Shalin, V. L. *Cognitive Task Analysis*. Erlbaum Associates, 2000.
22. Baker, R. S. J. d., Corbett, A. T. and Koedinger, K. R. The difficulty factors approach to the design of lessons in intelligent tutor curricula. *International Journal of Artificial Intelligence in Education, 17,* 4 (2007), 341-369.
23. Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. Example-tracing tutors: A new paradigm for intelligent tutoring systems. *International Journal of Artificial Intelligence in Education, 19,* 2, (2009), 105-154.
24. Charalambous, C. Y. and Pitta-Pantazi, D. Drawing on a Theoretical Model to Study Students' Understandings of Fractions. *Educational Studies in Mathematics*, *64,* 3 (2007), 293-316.
25. Beyer, H., & Holtzblatt, K. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann, 1998.
26. Große, C. S. and Renkl, A. Effects of multiple solution methods in mathematics learning. *Learning and Instruction*, *16* (2006), 122-138.
27. Yang, D.-C. and Reys, R. E. One fraction problem, many solution paths. *Mathematics Teaching in the Middle School, 7,* 3 (2001), 164-166.
28. Reed, J. A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13* (1987), 124-139.
29. Thußbas, C. Beeinflussen Cover Stories mathematischer Texaufgaben analogen Transfer in der Mathematik? *Zeitschrift für Experimentelle Psychologie*, *48*, 4 (2001), 272-289.
30. Catrambone, R. and Holyoak, K. J. Learning and subgoals and methods for solving probability problems. *Memory & Cognition*, *18* (1998), 593-603.
31. Thorndike, E. L. *The psychology of arithmetic*. Macmillan, 1922.
32. Pett, D. and Wilson, T. Color research and its application to the design of instructional materials. *Educational Technology Research and Development*, *44*, 3 (1996), 19-35.
33. Jakobsdottir, S., Krey, C. L. and Sales, G. C. Computer graphics: Preferences by gender in grades 2, 4, and 6. *Journal of Educational Research*, *88,* 2 (1994), 91-100.