

CycleTalk: Data Driven Design of Support for Simulation Based Learning

Carolyn P. Rosé, Rohit Kumar, Vincent Aleven, Allen Robinson, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh PA, 15213, USA
cprose@cs.cmu.edu, {rohitk,va0e,alr}@andrew.cmu.edu
http://www.cs.cmu.edu/~cprose

Chih Wu, US Naval Academy, 121 Blake Rd., Annapolis MD, 21402-5000, USA
wu@usna.edu

Abstract. In this article, we discuss the motivation for a novel style of tutorial dialogue system that emphasizes reflection in a simulation based exploratory learning environment called CyclePad, which was developed in order to offer students practice at design and optimization of thermodynamic cycles. We argue that while typical forms of exploration support offered in simulation based learning environments are meant to encourage the development of good exploration process skills, the instantiation of these typical forms of support in CyclePad are not sufficient, primarily because students don't choose to use them. We present a preliminary cognitive task analysis of design exploration tasks using CyclePad. Using this cognitive task analysis, we analyze data collected in two waves of formal data collection. Finally, we conclude with some system desiderata derived from our analysis as well as discussion of directions for continued experimental investigations related to tutor style.

Keywords. Tutorial dialogue, exploratory learning

INTRODUCTION

Tutorial dialogue is a unique, intensely dynamic form of instruction that can be highly adaptive to the individual needs of students (Nückles, Wittwer, & Renkl, 2003), and provides opportunities for students to make their thinking transparent to a tutor. The tutorial dialogue literature provides us with many convincing proofs of the technical feasibility of tutorial dialogue systems (Graesser, Bowers, Hacker, & Person, 1998; Rosé et al., 2001; Graesser, VanLehn, the TRG, & the NLT, 2002; Aleven, Koedinger, & Popescu, 2003; Evens, & Michael, 2006). What is needed now is more insight on how to wield that technology to the benefit of student learning beyond what is possible with more standard forms of interaction supported by typical state-of-the-art tutoring systems. Looking at naturalistic human tutorial dialogue inspires us to broaden our view of what intelligent tutoring systems can provide to students, and to consider forms of interaction that are not typically supported by current intelligent tutoring systems. One of the major research goals of the CycleTalk project (Rosé et al., 2004; Rosé et al., 2005) has been to

investigate the instructional effectiveness of novel ways of using tutorial dialogue technology in an exploratory learning environment.

Current tutorial dialogue systems focus on leading students through directed lines of reasoning to support conceptual understanding (Rosé et al., 2001), clarifying procedures (Zinn, Moore, & Core, 2002), or coaching the generation of explanations for justifying solutions (Vanlehn et al., 2002), problem solving steps (Aleven, Koedinger, & Popescu, 2003), predictions about complex systems (Evens & Michael, 2006), or computer literacy (Graesser, Moreno, & Marineau, 2003). Thus, to date, tutorial dialogue systems have primarily been used to support students in strongly directed types of task domains. We hypothesize that in the context of creative design activities, the adaptivity of dialogue will prove to have greater impact on learning than the impact that has been demonstrated in previous comparisons of tutorial dialogue to challenging alternative forms of instruction such as an otherwise equivalent targeted "mini-lesson" based approach (e.g., Graesser et al., 2002) or a "2nd-generation" intelligent tutoring system with simple support for self-explanation (e.g., Aleven, Popescue, & Koedinger, 2003).

We are conducting our research in the domain of thermodynamics, using as a foundation the CyclePad articulate simulator (Forbus et al., 1999). We predict that engaging students in natural language discussions about the pros and cons of their design choices is a highly interactive form of guided exploratory learning that is well suited to the purpose of science instruction. Such simulation environments have been used for science instruction for some time (e.g., D'Souza, Rickel, & Herreros, 2001), however previous systems have not been used to formally address research questions related to the relative effectiveness of tutorial dialogue instruction in comparison with alternative forms of exploration support. In the remainder of the paper, we present a preliminary cognitive task analysis of design exploration tasks using CyclePad. We then present an analysis of data collected in two initial studies of students using CyclePad, one in an unguided manner, and one in a Wizard of Oz scenario. We follow that by presenting evidence from the analysis of this data that suggests how tutorial dialogue can be used to assist students in their exploration process. We then discuss how we revised our materials based on findings from this initial wave of data collection, and follow up with a second wave of data collection. The analysis of this data suggests specific directions for system development and further experimentation.

SIMULATION BASED LEARNING AS A FORM OF EXPLORATORY LEARNING

In this paper we explore a particular form of exploratory learning involving discovery-based simulation environments (de Jong & van Joolingen, 1998). Discovery-based simulation environments require students to be actively involved, to plan, and to make inferences based on their observations (Swaak & de Jong, 2001). While simulation-based learning in particular, and exploratory learning in general, have been argued to have great promise for encouraging deep learning, the results from formal evaluations of their instructional effectiveness have often been disappointing (Mayer, 2004). Although the question of whether or not these environments encourage deeper learning of declarative knowledge is yet to be demonstrated conclusively, what is clear is that these environments offer students the opportunity to exercise their scientific discovery skills, which are important life skills. Thus, in this article we are specifically

concerned with supporting the exploratory process by encouraging the development of productive scientific discovery skills (Klahr & Dunbar, 1988; Reimann, 1991).

It is possible that discovery learning may lead to a different type of learning than more expository forms of instruction, such as practical or "intuitive" knowledge (Swaak & de Jong, 1996). For example, Carroll and colleagues (1985) found that students learned how to use a word processor better through exploration than through more expository means. Kashihara et al (2000) raise the issue that what type of learning is encouraged by the design of an exploratory environment might depend upon which type of cognitive load is reduced most as a result of the interaction design. For example, reducing the cognitive load for the exploration process may not result from the same or even compatible design features as reducing the cognitive load for the learning of the conceptually oriented material. They argue that in order to support a productive exploration process, exploratory design environments must be designed with the intention of reducing the cognitive load associated with exploration in order to encourage students to explore. In contrast, Swaak and de Jong (2001) provide evidence that the structure of the task may have more of an influence on student behavior than what is supported by the environment. In a comparison between two versions of an exploratory environment in which they manipulated how directive the system was, they did not find a difference in instructional effectiveness between two versions of the system. In both cases, students were presented with an appropriately ordered list of tasks. The only difference was whether they were required to go through them in order and to do each one. One explanation for the null result is that the task was sufficiently directive that the manipulation did not have enough of an effect on student behavior to make a difference in student learning. Striking the right balance between design desiderata at multiple levels is one factor making the development of effective exploratory learning environments difficult.

One important question to address is whether it is true that offering students more autonomy is the right way to encourage the development of productive scientific discovery skills. De Jong and van Joolingen (1998) argue that learners should have autonomy in their learning process rather than being led in a strongly directed manner by a tutoring system. They found that students were provided with sufficient structure just by being offered a suggested order for addressing problem solving goals based on difficulty of material rather than being required to complete all units or address them in the order presented. Swaak and de Jong (2001) explore the issue of how much control the system should have versus how much control the student should have. However, they do not provide a conclusive answer on what is the ideal delicate balance between student control and system control. In a similar spirit, White and Frederiksen (1990) have argued in favor of the idea of model progression where the system starts out with more control, limiting the search of the student, and progressively removes control over time. But again, this process has many degrees of freedom that are yet to be evaluated before specific design recommendations can be made with authority.

Beyond the difficulty of juggling multiple design concerns, it is challenging even to find a single, universal definition for what exploratory learning is. One popular conceptualization of exploratory learning is that what distinguishes "exploratory learning" from "non-exploratory learning" is the level at which goals are provided to the learner. Exploratory learning is associated with "high level goals" such as "survive in this simulation environment" or under specified goals such as "find all implications that can be drawn from these premises". Exploratory learning stands in contrast to more typical forms of instruction found in intelligent tutoring and computer based training systems such as (1) passive worked example studying

(Schworm & Renkl, 2006; Renkl, 2002), (2) active but totally guided tutorial learning (Smith, Hipp, & Biermann, 1995), (3) and most commonly, problem solving (Anderson et al., 1995; Koedinger et al., 1997). These forms of instruction are experimentally contrasted in (Charnay & Reder, 1986). On the macro-level, what is manipulated is the amount of structure provided for students. In (1) and (2), for example, students make no choices whatsoever, although students in (2) are more active than students in (1). High level goals are set, and low level steps are provided. In (3), problem solving goals are made for the student, but the student chooses how to satisfy those goals through means-ends analysis. In contrast, in exploratory learning, the student sets problem solving goals and chooses how to satisfy those goals. Thus, the student has greater autonomy, but the student is limited by their own conception of what is possible and valuable to explore. In (3) the student is prompted to explore areas in the space of possibilities that they may not have thought of by themselves. Furthermore, they reap the benefits of exploring alternative ways of achieving those goals. However, they do not get the practice at setting goals for themselves that students in exploratory learning environments get.

In depth analyses of student behavior with simulation environments have shown very convincingly that students lack the scientific discovery skills that are necessary to fully make use of what these environments have to offer (de Jong & van Joolingen, 1998). Because of this, the majority of current simulation based learning environments include support facilities to scaffold the student's discovery process. However, care must be taken on what type of support to offer. There is a concern that some forms of support offered to students in an exploratory environment impose additional cognitive load. In response to this concern, Kashihara et al. (2000) argue in favor of an Exploration Search Control approach in which a delicate balance is struck between system control and student control, although there is no empirical evaluation of this idea even in their more recent work (Kinshuk & Lin, 2003). In general, there are two types of support offered in simulation-based exploratory environments: access to background material, and scaffolding to structure the exploration process. As we will discuss in the next section, CyclePad already has the typical kind of support available in state-of-the-art simulation-based exploratory environments. Later we will argue, based on results of an observation of CyclePad in routine use, that students do not choose to make adequate use of the help features that are available in CyclePad. In contrast, we will offer examples of human tutors scaffolding good process habits with students and use this as the basis for arguing that tutorial dialogue has the potential to increase the use of support facilities built into exploratory environments and this increases their effectiveness.

LEARNING AND WORKING WITH CYCLEPAD

We are conducting our research in the domain of thermodynamics, using as a foundation the CyclePad articulate simulator (Forbus et al., 1999; Wu, 2002). CyclePad was developed with the intention of allowing students to engage in design activities earlier in their education than was possible previously. The goal is first to offer students practical experience with the application of domain principles, and second to reinforce and deepen their understanding of those principles. While the process of design is distinct from the scientific discovery process encouraged in other existing simulation based learning environments, it has much in common because of the emphasis on systematic exploration of a simulation space, analysis of collected data, and

planning actions based on this analysis. With CyclePad, students perform analyses on designs in order to build their own mental model of the relationships between cycle parameters. This understanding is meant to guide the design and optimization process. Thus, when students are using CyclePad fully as it was intended to be used, students are engaged in a process very much like that of scientific discovery learning. Nevertheless, while we observed students struggling with some of the same issues that are typical in simulation based exploratory learning, we also observed unique problems related to the process of creating and fully defining an initial state for a cycle that must be addressed in order to make interaction with CyclePad instructionally beneficial.

CyclePad has been in active use in a range of thermodynamics courses at the US Naval Academy and elsewhere since 1996 (Tuttle & Wu, 2001); a number of textbooks focus very strongly on activities with CyclePad (e.g., Wu, 2002). Active learning with CyclePad stands in stark contrast to traditional engineering instruction in thermodynamics, which emphasizes analysis rather than design. Qualitative evaluations of CyclePad have shown that students who use CyclePad have a deeper understanding of thermodynamics equations and a better handle on the meaning of technical terms (Baher, 1999).

In this section we offer an overview of the content that we intend for students to learn in their interactions with CyclePad, discuss a cognitive task analysis that outlines the process students are meant to engage in with CyclePad, and conclude with a discussion of the types of help facilities offered by CyclePad.

Subject Matter: Rankine Cycles

Our explorations of CyclePad use focus on design and optimization of thermodynamic cycles. A thermodynamic cycle processes energy by transforming a working fluid within a system of networked components (condensers, turbines, pumps, and such). Power plants, engines, and refrigerators are all examples of thermodynamic cycles. In particular, our explorations with CyclePad have focused on offering students the opportunity to explore the design and optimization of Rankine cycles. Rankine cycles are a type of heat engine that forms the foundation for the design of the majority of steam based power plants that create the majority of the electricity used in the United States. Design and optimization of Rankine cycles is an important topic for mechanical engineering students because even small differences in cycle efficiency can make the difference of millions of dollars a year in electricity across the United States. Furthermore, it is an intellectually challenging topic because it involves balancing many competing concerns such as the underlying physics and chemistry, economic factors, reliability and maintenance issues, and environmental impact issues.

There are three typical paradigms for design of Rankine cycles, namely the simple Rankine cycle, Rankine cycle with reheat, and Rankine cycle with regeneration. As students work with CyclePad on design and optimization of Rankine cycles, they start with these basic ideas and combine them into novel designs. While there are a huge number of possible designs, each approach to improving the efficiency of a Rankine cycle is based on the same three core ideas: (1) Increase the maximum temperature at which heat is added to the engine. (2) Decrease the minimum temperature at which heat is rejected from the engine. (3) Keep the quality of the steam entering the turbine high. In other words, the amount of moisture in the steam should be kept low in order to avoid damaging the blades of the turbines. One of the most important and

difficult lessons that students must learn is to see their explorations with CyclePad in these terms. Every effective optimization that they will apply will be based on one or more of these three core ideas.

To illustrate how these principles are applied to the optimization of a simple Rankine cycle, consider the following example optimizations:

- Adjusting the temperature and pressure of the fluid in the boiler is one possible modification. This is an instance of the idea of increasing the temperature at which heat is added to the cycle and will increase efficiency, up to the point where the materials cannot withstand the extreme conditions.
- Adding a reheat cycle reheats the working fluid before sending it through a second turbine. This requires extra energy to the second heater, but it is balanced by the work done by the second turbine. This is a way of decreasing the temperature at which heat is rejected from the cycle while keeping the quality of the steam high.
- Adding a regenerative cycle sends some of the steam leaving the turbine back to the water entering the boiler, which decreases the energy required to heat the water in the boiler. Again, this decreases the temperature at which heat is rejected from the cycle.

These modifications can be combined, and multiple stages of reheat and regeneration are often used to optimize efficiency, though the cost of additional parts must be weighed against the gains in efficiency.

Cognitive Task Analysis: Exploration in CyclePad

We have constructed a preliminary cognitive task analysis (See Figure 1) describing how students might use CyclePad in the type of scenario they encountered during these studies (i.e., to improve a simple Rankine cycle). Students begin with the topmost node in the task analysis diagram displayed in Figure 1, which corresponds to laying out the initial topology of a Rankine cycle using the widgets provided by CyclePad. For example, they may choose to construct the topology for a simple Rankine cycle, which consists of a heater, a turbine, a condenser, and a pump. Moving to the right in the diagram, students must next set values for key parameters associated with each widget until the cycle's state is fully defined. At that point, the student can explore the relationships between cycle parameters by doing what are called sensitivity analyses, which allow the student to observe how a dependent variable's value varies as an independent variable's value is manipulated. Students may experiment with a number of alternative designs. Based on their experience they can plan strategies for constructing cycle designs with higher efficiency. As part of this planning process, students may conduct more analyses and reflect upon their understanding of how thermodynamic cycles work. Once they have a plan for proceeding, they begin the cycle once more by creating a new cycle, and so on. Below we discuss each of these steps in greater detail.

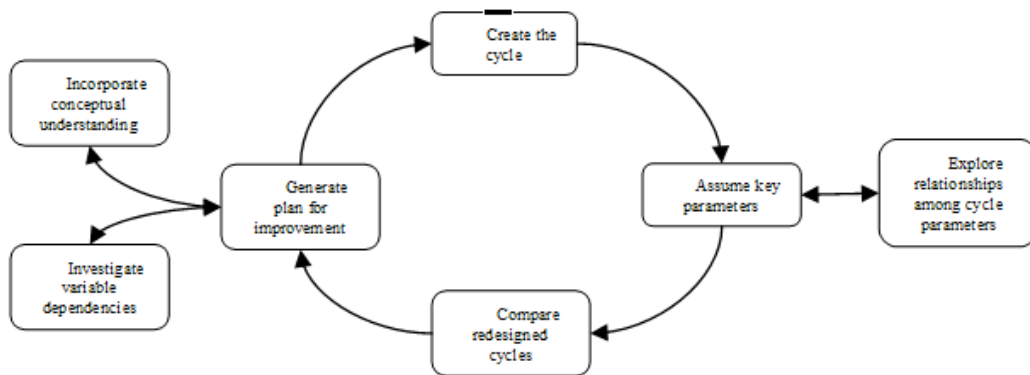


Fig.1. Cognitive task analysis.

- Creating the cycle and defining key parameters.* When creating a thermodynamic cycle according to the problem description, or modifying a given thermodynamic cycle, students must select and connect components to define a topology for their cycle. However, this is only the beginning of their design process. In addition to the topology of the cycle designs, each widget that is part of the topology has a number of parameters, some of whose values must be set by the designer or student, which is referred to as assuming values for cycle parameters. The CyclePad simulator will then compute as many additional parameters as can be derived from those assumptions that were explicitly set. When each parameter has a value, either given or inferred, CyclePad calculates the cycle's efficiency. In order to be successful, students must carefully select and connect components and be able to assume values in ways that acknowledge the relationships between the components. In fact, setting these values is the most difficult part of the process. Sometimes students inadvertently set inconsistent values between parameter values. When this occurs, a pop-up window labeled "There is a contradiction" alerts the student that this has occurred and that the student must address this problem before proceeding.
- Investigating Relationships Between Cycle Parameters.* Once the cycle state has been fully defined (i.e., the values of all parameters have been set or inferred), students can use CyclePad's sensitivity analysis tool to study the effect of possible modifications to these values. With this tool, students can plot the effect of manipulating the value of one variable on the value of another variable. These analyses may have implications for their redesign strategy. For example, when a Rankine cycle has been fully defined, students can plot the effect of the pressure of the output of the pump on the overall thermal efficiency of the cycle. The sensitivity analysis will show that up to a certain point, increasing the pressure will increase efficiency. The student can then adjust the pressure to its optimum level.
- Comparing Multiple Cycle Improvements.* Students can create their redesigned cycles, and, once the cycle states are fully defined, students can compute the improved cycle efficiency. Comparing cycle efficiencies of different redesigns lets students explore the problem space and achieve the highest efficiency possible. Suppose a student began

improving the efficiency of the Rankine cycle by including a regenerative cycle. It would then be possible to create an alternative design that included a reheat cycle (or several stages of reheat) and to compare the effects on efficiency before combining them. By comparing alternatives, the student has the potential to gain a deeper understanding of the design space and underlying thermodynamics principles and is likely to produce a better redesign.

CycleTalk Help Facilities

CyclePad offers the sort of help facilities that are typically found in simulation based learning environments (Forbus et al., 1999). For example, CyclePad offers access to needed background knowledge. In particular, students can get more information about parameters they need to set values for by clicking on them. When they do, the modeling assumptions that can legitimately be made about the selected component are displayed in English. CyclePad also has support for helping students resolve contradictions in the values they set for parameters when it detects them. It provides an interface that displays the contradicting settings along with some explanation about why they are incompatible. The student is free to select which setting to retract. A more process oriented type of support CyclePad offers the students is the sensitivity analysis tool, which allows students to view the relationships between cycle parameters. The interface guides the student to select an independent variable, a range for that variable, and a dependent variable. Coaching is also available on demand to aid the student in deciding upon a direction to go in their explorations. The on-board analysis coach uses a teleological analysis of the student's design to identify the likely functional roles of each component, which allows it to indicate problems in choices of parametric values based on these assumed roles. It can provide a limited form of instructional explanation on request.

CyclePad was designed with the intention of providing sufficient support for students to engage in a productive exploratory process (Forbus et al., 1999). However, in spite of its very impressive capabilities, it is plausible that CyclePad could be made even more effective. For example, CyclePad is geared towards explaining its inferences to students, at the student's request. It is likely to be more fruitful if the students do more of the explaining themselves, assisted by the system (Chi et al., 1981). Thus, a second area where CyclePad might be improved is in giving students the opportunity to develop their ability to think through their designs at a functional level and then explain and justify their designs as a form of reflection. A second way in which CyclePad's pedagogical approach may not be optimal is that students typically do not make effective use of on-demand help facilities offered by interactive learning environments (for a review of the relevant literature, see (Aleven et al., 2003)). That is, students using CyclePad may not necessarily seek out the information provided by the simulator, showing for example how the second law of thermodynamics applies to the cycle that they have built, with a possibly detrimental effect on their learning outcomes. Thus, students' experience with CyclePad may be enhanced if they were prompted at key points to reflect on how their conceptual knowledge relates to their design activities.

Below we discuss an investigation of routine use of the current version of CyclePad that demonstrates that CyclePad in its current form is not sufficient for encouraging a productive discovery process. One key finding we report is that the functionality that CyclePad offers for

supporting this process is not used or not used correctly by students, and thus does not have the opportunity to serve the purpose for which it was developed.

CYCLEPAD IN THE WILD AND IN THE LAB: NEEDS ASSESSMENT

We conducted an exploratory data collection effort to gather data to use as the foundation for the design of a support system for CyclePad that would be successful at helping students develop both good exploration process skills and deeper conceptual understanding of thermodynamic principles. In this section we discuss findings from two early data collection efforts that lay the ground work for the more extensive data collection and corpus analysis effort discussed in the subsequent section. We begin by describing the two early data collection efforts. We then discuss in detail observations about student difficulties with using CyclePad. We conclude the section by discussing observations about student difficulties with learning from their interactions with CyclePad. These observed difficulties form the foundation for a materials redesign and further data collection effort described in the subsequent section.

Data Collection

To begin our needs assessment, we collected two types of data: (1) data from naturalistic observation of CyclePad in typical use in a classroom setting, and (2) data from a small pilot study in which two forms of exploration support were contrasted.

The purpose of the first data collection effort was to observe unsupported use of CyclePad in order to determine where in the process of exploration with the current version of CyclePad students need the most support. This data from routine use of CyclePad was collected in the form of a take-home assignment administered to mechanical engineering students at the US Naval Academy. The professor of this course has used CyclePad in his courses for several years and is an expert CyclePad user. This professor's goal for CyclePad use is for his students to gain practical experience with applying thermodynamic principles in the design and optimization of thermodynamic cycles. As one of the regular homework assignments designed by this instructor, his class of 19 engineering students was asked to use CyclePad to improve the efficiency of a design for a shipboard version of a Rankine cycle. The students were given two weeks to complete the assignment. During this time, the students primarily worked independently with CyclePad. However, they did have access to their book and class notes, and to their instructor for occasional help. The data we collected consists of the students' written reports that they turned in, the written feedback and grades assigned by their instructor, and the log files of the students' interactions with the software.

The first data collection effort allowed us to get a realistic view of CyclePad use, but it did not provide us with the data we would need to explore issues related to learning with CyclePad. Furthermore, while it gave us a picture of the limitations of the current version of CyclePad, it did not offer us the opportunity to begin to evaluate our ideas for exploration support in connection with CyclePad use. Our long term vision has always been to design a new form of tutorial dialogue specifically for the purpose of exploration support. In order to begin to identify places in the process of CyclePad exploration where tutorial dialogue might have the greatest impact, we contrasted tutorial dialogue support from human tutors with completely non-

interactive support from a written script. The materials for this pilot study were developed in collaboration with a CMU mechanical engineering professor. His goals for eventual CyclePad use in his courses were more oriented towards deepening the students' conceptual understanding of thermodynamic principles. He designed a lab assignment that was meant to offer students the opportunity to work with the three main types of Rankine cycles described above, namely simple Rankine cycles, Rankine cycles with reheat, and Rankine cycles with regeneration. The written materials developed for this assignment, which we refer to as the Script, consist of some limited conceptual discussion about the three forms of Rankine cycles as well as instructions for building them. Suggestions for analyses to do with CyclePad as part of the process of optimizing the cycles is also included. As students worked through the lab, they were meant to spend time on each of the three types of Rankine cycles, building and optimizing each one in turn. A graduate student who is part of our team developed an introductory tutorial meant to introduce students to basic CyclePad functionality, but not discussing the domain content related to Rankine cycles in any depth. In close consultation with the professor and the book used in his course, we developed a pre/post-test to assess student learning. We designed isomorphic pre- and post-tests, each with 19 items covering background knowledge, basic thermodynamics concepts, and prediction problems. We counter-balanced the order of the pre- and post-test to control for possible differences in test difficulty. The experimental procedure for the pilot study was as follows:

- Pre-test (30 min)
- Introductory tutorial (30 min): Students worked through the tutorial and began to experiment with CyclePad independently.
- Experimental Manipulation (1 hour and 20 min): During this phase, students worked through the lab assignment, making as much progress as possible in the allotted time. In the experimental condition, students worked with a human tutor, whom they communicated with through typed chat. The tutor was able to view and interact with their CyclePad window through VNC. In the control condition, students had only the written materials, which the students in the experimental condition also had. We refer to these written materials as the Script. Students in both conditions had access to these written materials.
- Exploratory Design (20 min): During this phase, the written materials were taken away from students in both conditions, and they worked alone with CyclePad to build and optimize a Rankine cycle using what they learned.
- Post-test (30 min)

Two Carnegie Mellon mechanical engineering graduate students participated in the pilot study as tutors. We collected data from 4 students in the control condition and 5 in the experimental condition. We collected data from 3 additional students during the materials development phase of this small pilot study, providing additional example interactions with the human tutors for our development corpus, although these students did not strictly follow the experimental procedure mentioned above. Thus, we collected human tutoring protocols from 8 students altogether. Additionally, for all students participating in the pilot study, we have collected log files and screen movies of their interactions with CyclePad.

Student Difficulties Using CyclePad

Although the students from the US Naval Academy were experienced CyclePad users, and the Carnegie Mellon students had received training on how to use CyclePad, we observed students from both populations encountering a number of difficulties with using CyclePad, which we discuss here. In each case we offer an example of how we observed the human tutors assisting students in working through these types of difficulties.

Defining the Cycle State

As discussed above in connection with our cognitive task analysis, before a student can begin the optimization process with CyclePad, he must first build a topology out of the widgets provided by CyclePad and then fully define the internal state of the cycle by setting the values of the individual parameters associated with each widget, as described above. This is referred to as setting assumptions and is very difficult for students. This is a separate problem from those typically reported in connection with simulation based learning. It is not related to a lack of appropriate scientific discovery learning skills. If students start out without a concrete understanding of how cycle parameters relate to one another, they quickly find themselves in a catch-22 situation. They have trouble identifying which parameter values they must set, and what reasonable values would be. And CyclePad's most powerful facility for allowing students to explore relationships between cycle parameters, namely the sensitivity analysis tool, is not operational until the cycle's state is fully defined. Thus, the very functionality that would help them overcome their knowledge gaps is not accessible to them precisely because those knowledge gaps prevent them from getting far enough into the process to make use of it.

Despite CyclePad's built-in help functionality, we observed a number of students struggling when defining the state of each of the components in the cycle. On the take-home assignment assigned to the US Naval Academy students, 19 students were asked to improve the efficiency of a shipboard version of a Rankine cycle. The work of only 11 students resulted in a fully defined cycle such that the students were able to compute the efficiency of their improved cycle using CyclePad, even though these students had two weeks to complete the assignment and ample access to the professor. Of the 11 students who were able to fully define their improved cycle, 3 students created impractical or unworkable solutions and 3 other students did not improve the efficiency of the cycle in accordance with the problem statement. From the efforts of these students, we have seen that implementing one's redesign ideas in the CyclePad environment should not be considered trivial.

Because we only have the artifacts of these students' take-home assignments, it is difficult to speculate as to exactly what these students found difficult about fully defining the state of their improved cycle. However, we noted that on average, the redesigned cycles had 50% more components than the cycle that the students started out with. It seems likely that the greater complexity of the redesigned cycles that students constructed made it more difficult for students to identify the key parameters whose values must be assumed as part of the process of fully defining the cycle state. We have informally observed that our expert tutor is capable of defining the state of even complicated cycles in CyclePad without much, if any, trial and error. Presumably this is because he is intimately familiar with the relationships between cycle parameters, as opposed to novice students who may be struggling to maintain associations

between thermodynamic principles and their actions, especially as the number of components increases.

We did observe the complexity of defining the state of a redesigned cycle directly through observations of the Carnegie Mellon students who participated in our pilot study. In unguided work with CyclePad, we saw students having difficulty setting the assumptions for their improved cycle. One student was working for approximately 15 minutes on setting the parameters of a few components, but he encountered difficulty because he had not ordered the components in an appropriate way. The tutor was able to help him identify and remove the obstacle so that he could quickly make progress. When this tutoring episode began, the tutor asked the student to explain why he had set up the components in the non-functional order.

Student: I just figured I should put the exchanger before the htr

[The student is using "htr" to refer to the heater.]

Tutor: How do you think the heat exchanger performance/design will vary with the condition of the fluid flowing through it? What's the difference between the fluid going into the pump and flowing out of it?

Student: after the pump the water's at a high P

[P is an abbreviation for pressure.]

Tutor: Good! So how will that affect your heat exchanger design?

Student: if the exchanger is after the pump the heating shouldn't cause it to change phase because of the high pressure

...

Tutor: But why did you put a heat exchanger in?

Student: I was trying to make the cycle regenerative

...

Tutor: OK, making sure you didn't waste the energy flowing out of the turbine, right?

After the discussion with the tutor about the plan for the redesign, the student was able to make the proposed change to the cycle and define the improved cycle completely without any help from the tutor. Engaging in dialogue forces students to think through their redesign and catch errors that seem to be difficult for students to detect on their own. By initiating explanation about the design on a functional level, the tutor was able to elicit an expression of the student's thinking and give the student a greater chance for success in fully defining the improved cycle.

Investigating Relationships Between Cycle Parameters

One of the most useful tools that CyclePad offers students is the sensitivity analysis tool. A sensitivity analysis will plot the relationship between an independent variable (such as pressure or temperature) and a dependent variable (such as thermal efficiency). Information like this can

be very useful in planning one's approach to a redesign. It also provides a concrete way for students to observe the relationships between cycle parameters. Its very structure, requiring students to select one independent variable and one dependent variable, reinforces good scientific discovery process skills such as manipulating only one variable at a time. Nevertheless, the functionality only benefits students if they make use of it in a purposeful way.

There were two US Naval Academy students who performed large numbers of sensitivity analyses, as evidenced from their log files, but the comments from the professor on these students' written reports were critical of their process. They did not seem to document a well-reasoned path to their solution. From the relatively large numbers of sensitivity analyses in quick succession, one could speculate that these students' use of the sensitivity analysis tool was not purposeful. Rather, these students appeared to take a blanket approach in the hope that something useful might turn up. In contrast, in the CMU pilot study we observe the tutors assisting students to interpret sensitivity analyses and apply those interpretations to their designs in a systematic way, as illustrated in the following dialogue:

Student: I have recreated the basic cycle and am now in the sensitivity analysis

Tutor: Go ahead. Let's stick to eta-thermal

[student sets up the efficiency analysis in CyclePad]

Tutor: So what does this tell you?

Student: the higher the temperature to which the water is heated in the heater, the higher the thermal efficiency

Tutor: So do you want to try changing the peak temperature?

Student: to improve the efficiency, yes

In order to gain an understanding of how cycle parameters are related, which is crucial to fully defining the state of a cycle, students can ask CyclePad to explain the relationships between that parameter and other cycle parameters. Without guidance from a tutor, however, students tended not to take advantage of this facility, which is a finding that is consistent with results from other studies about help-seeking behavior in connection with instructional technology (Aleven et al., 2003). Investigation of the log files from the US Naval Academy take-home assignment reveals very limited use of CyclePad's explanation features. Only one log file indicated that a student had used the functionality more than ten times, and the log files of 8 of 19 students contained no evidence that the functionality was ever used¹. Similarly, in our direct observation of students working independently with CyclePad in our pilot study, we saw that students often set parameter values that contradict one another, causing errors that must be resolved before continuing. For example, one student encountered numerous contradictions on a single parameter over a short length of time, but still did not ask the system to explain how that parameter could be derived.

¹ We cannot rule out the possibility that students used the explanation features on CyclePad files they did not turn in or that they chose to ask their professor for help instead.

By contrast, students working with the tutor did seek out CyclePad's explanations, for example when the tutor asked them a question to which they could not respond. The tutor's prompts, and sometimes direct suggestions, encourage students to seek out and make use of CyclePad's help facilities.

Tutor: What does your efficiency depend on?

...

[Student asks CyclePad for "What equations mention eta-Carnot?"]

(eta-Carnot refers to the hypothetical efficiency of a completely reversible heat engine. The student is asking a theoretical question about how efficiency would be determined under ideal conditions.)

CyclePad displays $\eta\text{-Carnot}(\text{CYCLE}) = 1 - [T_{\min}(\text{CYCLE})/T_{\max}(\text{CYCLE})]$

Student: the definition of carnot efficiency

Tutor: What can you deduce from that?

Student: that the lower T_{\min}/T_{\max} , the higher the efficiency of the cycle; ie, the greater the temperature difference in the cycle the more efficient

Tutor: Is there any way you can modify your cycle accordingly?

The challenges faced by students working with CyclePad could be opportunities for learning. CyclePad has the capacity to explain how the cycles are functioning, but students do not seem to utilize CyclePad's articulate capacities spontaneously. When students are prompted to explain themselves and they receive feedback on their explanations, they are more likely to utilize CyclePad's helpful features in productive ways. Furthermore, as part of the discussion, the tutor may explicitly direct the student to seek out explanations from CyclePad.

Comparing Multiple Cycle Improvements

CyclePad makes it relatively easy for students to try alternative design ideas and thereby to generate high-quality designs. However, students working independently with CyclePad tended not to explore the breadth of the design space, even if they seemed to be aware of design ideas that would improve their design. Although students from the US Naval Academy observation who did the take-home assignment were exposed to both the reheat and regenerative strategies as part of their course prior to the assignment, only 8 of these 19 students incorporated both strategies into their redesigned cycles. Also, in the written report associated with the take-home assignment, the students were asked to explain the result of each strategy on the efficiency of the cycle. 15 of 19 students correctly explained why regeneration would improve the efficiency of the cycle. However, only 10 of 19 students used a regeneration strategy in their redesigned cycle.

In contrast, students working with the tutor in the Carnegie Mellon pilot study were prompted to consider many alternative approaches and were encouraged to contrast these alternatives with one another on the basis of materials and maintenance cost, in addition to cycle

efficiency. This explicit discussion of alternatives with the tutor is meant to lead to an optimal design. Here is an example dialogue where the tutor is leading a student to consider alternative possibilities:

Tutor: Yes, very good. How do you think you can make it better? i.e. how will you optimize the new component?

Student: we could heat up the water more

Tutor: That's one, try it out. What do you learn?

Student: the efficiency increases pretty steadily with the increased heating - should i put the materials limitation on like there was earlier? or are we not considering that right now

Tutor: OK, how about other parameters? Obviously this temperature effect is something to keep in mind. Include the material effect when you start modifying the cycle

Student: ok

Tutor: What else can you change?

Student: pump pressure

Tutor: So what does the sensitivity plot with respect to pump pressure tell you?

Student: so there's kind of a practical limit to increasing pump pressure, after a while there's not much benefit to it

Tutor: Good. What other parameters can you change?

Student: exit state of the turbine

Tutor: Only pressure appears to be changeable, let's do it. What's your operating range?

Student: 100 to 15000. right?

Tutor: Do you want to try another range? Or does this plot suggest something?

Student: we could reject even lower, since its a closed cycle

Tutor: Good!

Student Difficulties Learning with CyclePad

Beyond the difficulties students have on individual steps of the exploration process, students have high level problems navigating the design space in a purposeful, principle driven way that would allow them to learn effectively from their interaction with the system. Students often get lost in the process of setting this large number of parameters and lose sight of the three basic ideas that should guide their exploration, although reinforcing those principles is of utmost

importance to the learning objectives of the assignment. Finding it a struggle to maintain a view beneath the surface details to find ways of purposely using the three basic principles to guide their search, they may fall into a pattern of more or less undirected tweaking, which is unproductive both for their performance at optimizing the cycle as well as for their learning. One specific consequence of the lack of ability to see the exploration process in terms of deep principles is that students easily fall into the trap of believing that more sophisticated designs will be more efficient. Thus, students working unsupported with CyclePad have a tendency to be drawn towards the more complex portions of the design space before they are ready to fully understand how to use that sophistication to an efficiency advantage. We have observed that when our tutors observe this behavior, they encourage students to keep it simple and direct them back to more basic design explorations until students demonstrate a solid understanding at that basic level. Thus, one of the important functions of the tutors is to regulate the pace of the students as they move through the material. The tutors appear to be guided by a policy of maximizing the instructional advantage of student exploration. They do not discourage students from attempting design modifications that will not ultimately yield an optimal design, but they do discourage students from explorations that they are not prepared to learn from. This high level structuring provides many advantages for students. Because of it, students are not hampered by their preconceptions that would have led them to spend their time in explorations that would have been devoid of educational value. Nevertheless the students still play an integral role in deciding how they will spend their exploratory time. The give and take of this human tutoring interaction allows students to reap the best of both worlds – with the structure of problem solving but the practice at high level decision making and motivational benefits thereof associated with exploratory learning.

We expected that interacting with a human tutor while working with CyclePad would make the experience more beneficial for learning. We observed the tutors successfully helping students to move through the cycle creation and optimization process. As the tutors worked through the assignment with the students, they required the students to think through choices and then act according to those choices. Before asking the students to make a choice, the tutors tended to engage the students in a discussion about the options available, often encouraging the students to take the lead in enumerating the set of options. We observed students in the experimental condition behaving at a high level of activity throughout their interaction with CyclePad and the tutor. In particular, students frequently asked questions. For one student the average was 23 questions per hour. In total, 12.2% of student contributions consisted entirely of student initiatives, including unsolicited questions, observations, and arguments. Students also frequently offered unsolicited explanations as elaborations on their answers to tutor questions. One consequence of the intense interaction between the tutors and the students was that exploration supported by a tutor took more time than exploration guided by the Script. Because of this there were differences in how students in the two conditions explored the design space. In general, students in the script condition covered the entire Script, whereas students working with a human tutor mainly covered the simple Rankine cycle in the same period of time. Observations from the assessments used in the pilot study revealed areas of concern in connection with student learning with CyclePad in both conditions, although the tutors behaved in a way that was consistent with the goal of enhancing student conceptual understanding.

With such a small number of data points, we did not expect to see statistically significant differences between conditions on the test and we did not find any. Nevertheless, there was a

significant gain between pre-test and post-test in both conditions overall. However, a troublesome observation was that there was no significant difference between pre-test and post-test specifically on the prediction problems where students were asked to predict what the effect on one parameter (such as efficiency) would be if another parameter was manipulated (such as minimum pressure, or P_{min}). In fact, the average post-test score on questions related to relationships between cycle parameters was slightly lower than that of the pre-test. Understanding the relationships between cycle parameters is at the core of a deep conceptual understanding of Rankine cycles. While these results are preliminary due to the very small population size, we were concerned about the lack of gain on prediction problems since we expected the ability to make such predictions to follow from experience working through the process of optimizing cycles in CyclePad. In order to address this concern in subsequent iterations of materials development, as we will discuss below, we placed a greater explicit emphasis on intentional reflection on the relationships between parameters in the cycles.

Another concern was in connection with the results from free exploration with CyclePad. 2 out of 4 Control condition students and 1 out of 5 experimental condition students were not able to fully define a Rankine cycle without support during the free exploration assessment. Thus, while students worked through fully defining and then optimizing Rankine cycles during the hour and 50 minutes prior to the free exploration phase, they did not master the ability to fully define a cycle on their own. Of the students who did manage to build and at least partly optimize a cycle, the average efficiency of the final cycles were only a modest increase over the minimum of 37.4. The average in the Control condition was 39.2, whereas it was 42.9 in the Experimental condition. In general, students in the Control condition attempted more complex designs, which might explain the slight difference in success rate observed. Students in the script condition who produced an analyzable design all chose to implement a combined reheat and regeneration cycle while the students in the tutoring condition universally chose to implement a cycle with one or more stages of reheat. A similar pattern of difficulty with fully defining Rankine cycles was observed with the US Naval Academy students as already discussed.

Discussion

At the end of our first round of exploratory data collection, we reflected on what we learned that would inform the development of the CycleTalk tutorial dialogue system. One important lesson learned was that, while consistent with the literature on simulation based learning, we did observe students lacking in the area of scientific discovery skills. This was not the biggest hurdle that was standing in the way of students succeeding either with the practical objective of using CyclePad to build and optimize a cycle or with the learning objective of imparting to students a deep understanding of the important principles that underlie the design of thermodynamic cycles. Our target student population does not start out with a practical knowledge of the three principles that would guide them in purposeful exploration with CyclePad. We learned that the biggest gap in student learning with CyclePad is related to learning the relationships between cycle parameters. Furthermore, lack of knowledge in this area is one important reason why students get stuck at an early stage in the process of building, analyzing, and optimizing a cycle with CyclePad. In our pilot study at Carnegie Mellon we observed that the support our human tutors offered students when they got stuck, and to some extent also the Script, was sufficient scaffolding for helping the students move past the hurdles that would have otherwise stymied

their progress early in the process, nevertheless it was not enough to allow students to internalize the relationships between cycle parameters that they were meant to learn from the experience.

INVESTIGATING THE SUPPORT OF A DEEPER UNDERSTANDING OF RANKINE CYCLES

In response to the findings from our first round of data collection discussed in the previous section, we made a number of changes in our approach. Most importantly, we increased our emphasis on conceptual understanding of thermodynamics. The Carnegie Mellon University mechanical engineering professor on our team added a substantial introductory write-up about Rankine cycles, explaining the three key principles and how they relate to the three main forms of Rankine cycles, namely simple Rankine cycles, Rankine cycles with reheat, and Rankine cycles with regeneration. He also revised the Script, adding in more conceptual discussion along the way and expanding the number of suggested sensitivity analyses to encourage more exploration of relationships between cycle parameters. Small revisions were also made to the instructions for setting up the cycle and defining the initial cycle state to scaffold the students in moving past this difficult hurdle so they can spend more time exploring and hopefully internalizing the relationships between cycle parameters. Additional small revisions were made to the materials used for measuring learning with CyclePad, as discussed below. At the end of the Fall semester of 2005 we conducted another data collection effort in the context of a sophomore thermodynamics course at Carnegie Mellon University. 23 students from the course worked through our revised materials with the aid of one of three human tutors. In this section we describe that data collection effort and what we learned from it. While we do not evaluate our success in comparison to a control condition here, as a validation of our revised materials and approach we demonstrate a significant knowledge gain on the topic of relationships between cycle parameters across the student population, and a 100% success rate for one of our tutors at imparting the ability to build and fully define a cycle as assessed by the free exploration practical assessment.

Corpus Collection and Learning Assessment

We collected our corpus over a two week period of time as part of a sophomore thermodynamics course at Carnegie Mellon University beginning the week when Rankine cycles were introduced in the lecture portion of their class. 23 students were tutored by one of three mechanical engineering graduate students during an individual tutoring session². The study consisted of two labs involving work with CyclePad. The first lab was a self-paced take-home assignment done during the first week of the study. The second lab was a 3-hour on-campus lab session

² This data collection effort was part of a larger 3X2 factorial design, with main results reported in (Rosé et al., 2005) and further analysis in (Arguello et al., to appear). The human tutoring condition reported here was one of three goal level conditions in the larger study. These three conditions were crossed with two goal orientation conditions. Thus, the students whose data we discuss here were randomly assigned to two different goal orientation conditions.

completed during the second week of the study. We strictly controlled for time. The 3-hour lab session was divided into 7 segments:

- After completing the consent form, students were given 20 minutes to work through a 50 point pre-test consisting of short answer and multiple choice questions covering basic concepts related to Rankine cycles, with a heavy emphasis on understanding dependencies between cycle parameters.
- Students then spent 15 minutes reading an 11 page overview of basic concepts of Rankine cycles.
- Next they spent 25 minutes working through the first of three focused materials with readings related to simple Rankine cycles, suggested problem solving goals, and analyses to help in meeting those goals.
- Next they spent 20 minutes working through the second set of focused materials related to Rankine cycles with reheat.
- They then spent 20 minutes working through the third set of focused materials related to Rankine cycles with regeneration.
- They then spent 40 minutes in a Free Exploration phase creating the most efficient Rankine cycle they could with no instructional support either from the tutor or any of the written instructional materials they had previously been given.
- They then spent 20 minutes taking a post-test that was identical to the pre-test.

The domain specific materials used in the study, which consisted of a take-home assignment, pre/post-test, introductory reading material about Rankine cycles, and focused readings with suggested illustrative analyses to perform using the CyclePad simulator for three forms of Rankine cycles, were all developed by a Carnegie Mellon University mechanical engineering professor with the help of three of his graduate students and minimal input from our team. Similar to the pilot study described above, we collected logs of all student behavior with CyclePad as well as logs of their conversations with the tutors. Students and tutors were in separate rooms across campus from one another. However, the tutors were able to observe the students' behavior with CyclePad through VNC and were able to communicate with the students through typed chat.

Similar to the pilot study described in the previous section, we measured learning by means of a two part pre/post-test and the results of the free exploration segment. 32 multiple choice and short answer questions were used to test analytical knowledge of Rankine cycles, including relationships between cycle parameters. An important aspect of this was a set of prediction questions where students were told to predict the impact of a specific change in one cycle parameter on several other cycle parameters. The other part of the test was a set of 9 open response questions assessing conceptual understanding of Rankine cycles. An analysis of the test results as well as performance on the free exploration practical assessment demonstrated that we were successful this time at increasing student conceptual knowledge of thermodynamics principles, understanding of the relationships between cycle parameters, and practical ability to build and optimize cycles. Overall, there was a significant increase in test performance between pre-test and post-test ($F(1, 45) = 23.18, p < .001$, effect size 1.4 standard deviations).

Furthermore, there was a significant increase within both the conceptual portion of the test ($F(1,45) = 21.81, p < .001$, effect size 1.3 standard deviations) as well as the analytical portion ($F(1,45) = 11.47, p < .001$, effect size 1 standard deviation), which specifically focuses on the relationship between cycle parameters. On the free exploration practical assessment, there was a significant difference in effectiveness between tutors (as measured by a binary logistic regression, $p < .001$). 100% of the students tutored by one of the three tutors were successful at building and fully defining a cycle, whereas 0% of those tutored by another of the three tutors were successful, and 38% of those of the final tutor were successful.

Corpus Analysis

Using an exploratory corpus analysis tool called InfoMagnets (Arguello et al., to appear), a thermodynamics domain expert constructed a topic analysis of the corpus of dialogues collected during the study. The InfoMagnets interface facilitates the process of exploratory analysis of corpus data. Using a measure of lexical cohesion, InfoMagnets takes as input a collection of contextually-related sample human-human dialogues, segments each based on its estimate of topic boundaries, and automatically clusters these dialogue segments into topically-related groups. Once the segments are automatically assigned to classes, the user explores the corpus using its interactive corpus visualization interface and optionally adjusts the segment boundaries and document-to-category assignments such that the resulting organization is more in tune with his or her own mental model of the domain content.

It took the expert two full working days to complete the analysis of dialogues from 21 students³, with 65 minutes worth of conversational data from each student. Altogether each student's protocol was divided into between 10 and 25 segments such that the entire corpus was divided into approximately 379 topic segments altogether. The resulting analysis consisted of 15 distinct topics, with each student covering between 4 and 11 of these topics either once or multiple times throughout their interaction with the tutor. In support of the instructional value of these discussion topics, we found a strong and reliable correlation between topic coverage (i.e., percentage of topics discussed at least once) and post-test score, with pre-test score used as a covariate ($R^2 = .715, N=21, p < .05$). The topic analysis of the corpus gives us a way of quickly getting a sense of how tutors divided their instructional time between different topics of conversation. For example, we can quickly get a sense of how consistently tutors were in the instruction they imparted to students. Specifically, in our corpus we observed a great deal of variability in topic coverage across students. In particular, the average correlation between the topic distribution for an individual student and that of the corpus as a whole was low (average $R^2 = .29, N=21$). Thus, we could suspect that our tutors were very adept at adapting their material to the individual needs of their students. However, we did not find a significant correlation between the distribution of topics students performed either relatively well on or relatively poorly on in the pre-test and those covered during the tutoring interactions. Nevertheless, the topic labeled corpus provides us with an easy way of retrieving alternative examples of how our tutors approached these topics with students, which we can then pattern our authored dialogue system after.

³ Unfortunately, the protocols of two out of the 23 students were lost and thus not included in this analysis.

The 15 topics identified in the corpus by the domain expert fall roughly into 4 clusters. The first cluster, which we refer to as *social talk*, consists of two topics that are primarily social in nature, specifically greetings and off-topic conversation. The next three topics, which we refer to as *cycle construction and maintenance topics*, focus on the mechanics of building the initial cycle or modifying the cycle based on the result of efficiency analyses. 6 topics related to basic thermodynamics concepts related to Rankine cycles are classified as *understanding foundational concepts*. A final set of topics cover the concept of sensitivity analyses and several specific sensitivity analyses students did as part of their instruction. We refer to this final set as *exploring relationships between parameters*. See Figure 2 for an overview. In the remainder of this section, we discuss each of these topic clusters in turn. To give an estimate of what tutors emphasized and how they spent their time with the students, we indicate in the tables within each subsection the number of instances of each topic in our corpus.

Social Talk

Nearly 80% of the topic segments in the corpus were directly related to thermodynamics. However, there was a noticeable amount of social talk interspersed with the domain focused segments. For example, the protocols between students and tutors typically began with social talk, most often a greeting, in which the student and tutor established a rapport. Other off-topic conversation was labeled as General Talk. This frequently included general suggestions about learning strategies or issues related to managing the coordination between tutor and student. An overview of these topics is found in Table 1.

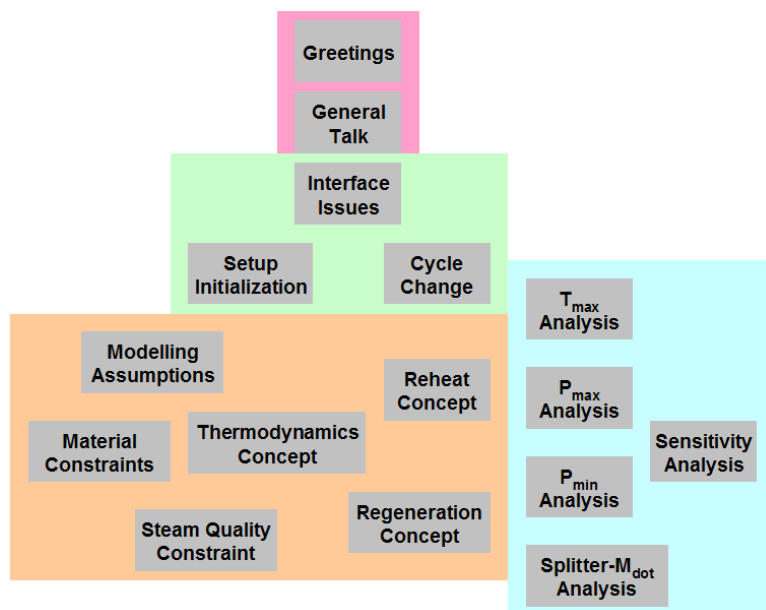


Fig.2. Topic analysis constructed using InfoMagnets exploratory corpus analysis tool, which is described in (Arguello et al., to appear).

Table 1
Topic segments related to Social Talk

Topic	#Instances	Description
Greetings	22	Segments containing conversational openings and closings, such as "Hello" and "Goodbye"
General Talk	56	Segments in which the tutor and student are engaged in discussion not specifically related to thermodynamics, such as when the tutor is giving the student a "pep talk".

Here is an example segment assigned to the category of General Talk. In this segment the tutor is admonishing the student for arguing with his advice.

Tutor: the only way a person can learn is to be teachable. i'm your friend ready to help. ok?

Student: i know. i thought i opened s6, and i was wrong. Sorry.

Tutor: don't waste time arguing.

In the future we plan to do a detailed analysis of the connection between social talk in our corpus and student learning. We have informally observed that the tutor who was least successful at imparting both practical and theoretical knowledge to his students was also much less successful at establishing a positive rapport with his students. Not only was it the case that 0% of his students were able to produce a fully defined cycle during the free exploration, but they gained significantly less knowledge between pre- and post-test than those tutored by the other two tutors. However, beyond his lack of relational success with his students, his tutoring style was strikingly different along other dimensions as well, which may be responsible for the difference in effectiveness. For example, he frequently lost patience with students and took over the problem solving rather than letting the students proceed as independently as possible with his support.

Cycle Construction and Maintenance

Approximately 27% of the topic segments in the corpus were related to the nuts and bolts of building and modifying cycles using CyclePad. This includes segments related to the initial construction of a fully defined cycle to begin the optimization process as well as discussion related to modifying the cycle after sensitivity analyses are used to learn about relationships between cycle parameters. Because a large portion of student struggles with the CyclePad interface occurred in connection with these activities, we have also included segments related to all types of interface issues with CyclePad in this group. This group does not include either conceptual discussion about principles underlying the design of cycles nor discussion related to evaluating the relationships between cycle parameters using the sensitivity analysis tool. These other two groups of topics are discussed in the two subsections immediately following this one. An overview of cycle construction and maintenance topics is found in Table 2.

Table 2
Topics related to cycle construction and maintenance

Topic	#Instances	Description
Setup/ Initialization	41	Segments where the student and tutors are talking about constructing an initial cycle to begin their investigation. Segments about modeling assumptions to be made for initialization do not count under this topic since this discussion was considered conceptual in nature. Dialogue about parameter values used to define the initial state of the cycle fall under this topic.
Cycle Change	31	Segments where a cycle is being modified after some analysis are included in this category.
Interface Issues	31	Any segment where the issue of how to accomplish something with the CyclePad interface is being discussed between the student and tutor

Understanding Foundational Concepts

Roughly 26% of the topic segments were focused on conceptual discussion. This includes general conceptual discussion related to all Rankine cycles as well as that pertaining specifically to either Rankine cycles with reheat or Rankine cycles with regeneration. Table 3 offers an overview of the types of conceptual discussion found in the protocols.

Table 3
Topic segments targeting foundational concepts related to Rankine cycles

Topic	#Instances	Description
Thermodynamics Concepts	37	Segments in which the student and tutor are talking about a wide range of elementary thermodynamic concepts that are not specific to reheat or regeneration principles. For example dialogues related to types of thermodynamic processes, Carnot cycle and its efficiency, and the laws of thermodynamics were included here.
Reheat Concepts	10	Segments in which the student and tutor are talking about use of reheat in Rankine cycles and why it improves efficiency. This cluster also includes segments about how reheat is implemented in practice and particular issues related to defining the state of a cycle with reheat.
Regeneration Concepts	10	Analogous to reheat concepts, except pertaining to regeneration instead.
Modeling Assumptions	21	Segments in which the tutor and the student are talking about setting the modeling assumptions of the cycle, related to the idealized Rankine cycle.
Steam Quality Constraint	13	Segments that talk about the need to maintain the steam quality above a specified limit at the exit of turbine and the consequences of violating this constraint.
Material Constraints	8	Discussion about design decisions related to limitations in the ability of certain materials to withstand temperatures and pressures beyond certain limits.

The dialogue afforded the tutors the opportunity to tie in discussion about foundational topics with practical applications of those topics. For example, in the segment below, the tutor discusses the motivation for considering material constraints and then leads the student to select a material and then investigate its own specific constraints related to maximum temperature.

Tutor: yes but we are constrained by material of boiler. we cannot go above 570 c. right?

Student: ok...did it tell us that somewhere? i missed that.

Tutor: well go ahead and choose stainless steel as material of heater first

Tutor: ok now try changing the t of s2 as 600 c

Student: ok

Tutor: it says the max limit for stainless steel was 576.9 c. so we limit it below that t. why do you think high boiler t increases efficiency?

Student: so should i put the temp at s2 to 576.9?

Tutor: ok...see, it doesnt allow that!

Student: ahh but it allows .8

Tutor: it has some safety factor

Exploring Relationships Between Parameters

The final 26% of topic segments are related to the important issue of exploring relationships between cycle parameters. Because this was a topic of concern from our preliminary data collection effort described in the previous section – both with respect to pre- and post-test gains and results from the free exploration practical assessment, we give this group of topic segments particular consideration here.

The first cluster of these segments, labeled Sensitivity Analysis, relates to high level discussion about doing sensitivity analyses or segments in which student and tutor were working through sensitivity analyses together, but it was not clear from the discussion which sensitivity analysis they were working on. The other categories of segments that fall into this set were each related to a specific sensitivity analysis. See Table 4 for an overview of topics related to exploring relationships between cycle parameters.

The results from our first pilot study indicate that gaining an understanding of the relationships between cycle parameters does not automatically come from going through the process of building and optimizing cycles. Thus, we can not consider the answer to effectively impart knowing in this area to be as simple as merely scaffolding the process of performing as many sensitivity analyses as possible. While we found a correlation between number of sensitivity analyses accomplished by students and learning on the pre/post-test, it was not the case that students who did an above median number of sensitivity analyses learned significantly more than those who did a below the median number of sensitivity analyses. Furthermore, the

most effective tutor did not lead his students to do significantly more sensitivity analyses than the other tutors. In our continued corpus analysis work we are exploring differences in how the three tutors used sensitivity analyses as opportunities to stimulate reflection.

Table 4
Topic segments related to exploring the relationships between cycle parameters using sensitivity analyses

Topic	#Instances	Description
Sensitivity Analysis	10	Segments where the tutor and student are talking about performing sensitivity analysis on some of the parameters but when it is not clear which sensitivity analysis is being performed. Also segments when multiple sensitivity analyses are performed without much dialogue about the analysis.
Tmax	27	Segments in which the student and tutor are talking about the sensitivity analysis for seeing the effects of maximum temperature on efficiency and/or steam quality. Includes dialogue about interpretation of the analysis, setting parameters for the analysis and conclusions from the analysis.
Pmax	17	Similar to Tmax analysis, but related to sensitivity analysis of maximum pressure in the cycle
Pmin	22	Same as Tmax and Pmax analysis, but related to sensitivity analysis of minimum pressure in the cycle.
Splitter-Mdot	23	Similar to above, but dialogue about sensitivity analysis to see the effect of changing splitting fraction in regenerative Rankine cycle.

The tutors did not simply scaffold the process of doing sensitivity analyses. They used them as an opportunity for the students to reflect on the relationships between cycle parameters. Nevertheless, the two tutors who were successful in terms of knowledge gain for their students on the pre/post-test used markedly different strategies for leading students through sensitivity analyses. The tutor whose students were 100% successful at creating a fully defined cycle during the free exploration practical assessment mainly restricted his comments to offering the student exploration goals, although he answered student questions when asked and occasionally presented students with explicit reflection questions. Here is an example dialogue with him leading a student to experiment with a series of sensitivity analyses.

Tutor: do the sensitivity. you are doing good

Student: ok

Tutor: try other parametrs as well...maybe they will have more effect..tmax and mdot

Student: i already played with mdot

Tutor: ok...also dec pmin helps...try other things

Student: p in s5...that's as low as it can be...the quality is .8507

Tutor: ok...try tmax

Student: 575

Tutor: why won't it plot mdot higher than .5 for s10?

The other tutor who was successful in terms of his students' pre/post-test knowledge gains but not as much on the free exploration practical assessment focused more on conceptual discussion. Once that tutor helped a student do the analysis, the tutor would often ask the student to interpret the graph and explain the whys and hows. He would then encourage the student to respond by making a corresponding adjustment to the cycle. Below is such a dialogue connected with a sensitivity analysis exploring the relationship between Tmax and cycle efficiency.

Tutor: so lets do sensitivity analyses. well we need to plot cycle efficiency. And you need to change the range.

Student: ?

Tutor: refer page 5. the suggested range for t odf s2 is 300-575 c

Tutor: ok. so does this plot make sense to you?

Student: yes

Tutor: why?

Student: so i should put temp at the highest it can be because this will increase efficiency as much as possible. so 500

Because the first tutor was successful both in terms of the test and in terms of the practical assessment, we must conclude that his strategy was more appropriate overall. He did not lead his students to conduct significantly more sensitivity analyses than the other effective tutor. However, because his directives were kept at a high level, the burden was on his students to make choices and observe the effects of those choices. He asked his students significantly fewer explicit reflection questions than the other effective tutor ($F(2,19)=15.83$, $p < .001$, effect size for pairwise comparison 2.7 standard deviations, computed from Tukey posthoc analysis). Nevertheless, by placing the burden of choice on his students, he may have effectively stimulated the same reflection as part of the process. Comments such as "maybe they will have more effect" might remind the student to think about the relationships between parameters while making choices. One potential explanation is that he was more successful than the second tutor because his approach taught his students to explore in a reflective manner, rather than merely reflecting after exploring. Further experimental investigation is required to test this hypothesis.

Discussion

The results of the corpus analysis from our Fall 2005 data collection effort do not definitively answer the question of how to build the most effective tutorial dialogue system for supporting exploration in a simulation environment. However, they do lead to interesting research questions

to investigate through further experimental work as well as offering some starting points for our system development effort. One obvious set of questions would focus on teasing apart the impact of the marked stylistic differences between the two tutors who demonstrated their tutoring prowess in terms of their students' pre- to post-test knowledge gains. One specific point of inquiry under that heading might be a further exploration of the instructional benefit of why questions (Rosé et al., 2003). We reported above a large and statistically reliable difference in the occurrence of why questions and other reflection oriented questions between our two effective tutors. We would like to probe deeper than the differences in surface characteristics of their respective styles to see if we might find a level of abstraction on which we can find commonalities between their respective approaches for eliciting reflection from their students.

Another set of important questions relate to evaluating the instructional value of exploration for learning. In the data collection effort discussed in this section, students were told that they were free to explore the instructional materials in any order they saw fit in terms of meeting their own instructional goals. We reported a low correlation on average between overall distribution of topic segments and those of each individual student, which could be indicative of tutor adaptation to the individual needs of students, although no predictable pattern of topic coverage related to prior knowledge as assessed by the pre-test could be discerned. It is possible that students might benefit more from instruction that specifically emphasized the topics they demonstrated weakness on in the pre-test. We cannot assume that the content as presented in this corpus is adapted to the needs of the students in an optimal way.

Another direction for future investigation is the social aspects of the interaction between tutor and student. We discussed above that our least successful tutor was also our least socially adept tutor. This raises questions related to the importance of social intelligence in connection with tutorial dialogue agents. One hypothesis is that student receptivity is directly related to the extent to which tutors establish and maintain a positive rapport with their students. While the corpus analysis presented here provides only anecdotal evidence in support of this hypothesis, some other recent experimental work does support this hypothesis, especially in connection with tutor politeness (Wang et al., 2005). Beyond the issue of politeness, however, the corpus analysis presented here raises new questions. For example, we observed a large number of topic segments in our corpus related to interaction management between tutor and student. This leads us to ask to what extent conversation related to interaction management influences the shape of the interaction between student and tutor and how this influence, if any, affects the instructional effectiveness of the interaction.

These important remaining questions relating to how best to support learning in a simulation environment are related mainly to tutoring strategy and style rather than content. The topic analysis of the corpus provides us with valuable data offering specific directions for content development. We know these are topics that come up frequently as students are working through the materials. Furthermore, we can begin to experimentally evaluate some of our hypotheses by implementing dialogue agents using the same content identified through our topic analysis, but instantiated by means of different dialogue strategies according to the research questions enumerated above. In our previous work we have explored a hybrid dialogue agent - wizard-of-oz setups that can be implemented efficiently and are effective for maintaining consistency of content across conditions and style within conditions (Rosé & Torrey, 2005). Building upon our prior work on knowledge construction dialogues (Rosé & VanLehn, 2005; Rosé et al., 2001; Jordan, Rosé & VanLehn, 2001), we have begun to develop a dialogue authoring environment

called TuTalk that facilitates the process of authoring dialogue agents from corpus data (Gweon et al., 2005). Thus, an important direction for our current work is to use this infrastructure to begin to address these important questions related to tutor style.

CONCLUSIONS AND CURRENT DIRECTIONS

In this article, we have discussed the motivation from two waves of data collection and analysis for a novel style of tutorial dialogue system that emphasizes reflection in simulation based exploratory learning environments.

We first presented a preliminary cognitive task analysis of design exploration tasks using the simulation based learning environment called CyclePad. Using this cognitive task analysis, we presented an analysis of data collected in an initial wave of data collection in the form of two small exploratory studies of students using CyclePad, one in an unguided manner, and one in a Wizard of Oz scenario involving human tutors. Using data from unguided CyclePad use, we argued that students do not choose to make use of the help facilities that CyclePad offers. Instead, they fall into a pattern of shallow tweaking, which does not lead to meaningful learning experiences. Using data from guided CyclePad use where students had the support of a human tutor through text-based chat interactions, we have argued that tutorial dialogue can be used to guide students along productive exploration paths and encourage them to reflect and learn from their interactions with CyclePad. Thus, this analysis suggests ways in which tutorial dialogue can be used to assist students in their exploration and encourage a fruitful learning orientation. However, while the tutors were successful in assisting students in working with CyclePad, we identified a lack of success with imparting an understanding of the relationship between cycle parameters and the ability to build and fully define a cycle using CyclePad without support.

Following discussion of our first wave of data collection, we then discussed a second wave of data collection, which took place after we modified our instructional materials based on findings from the first wave of data collection. The results from this investigation showed more promise in terms of student learning as measured by pre to post-test gains as well as success (at least in connection with one of three tutors) at imparting the ability to create and fully define a cycle independently. We concluded by presenting a topic analysis of the collected corpus as well as a discussion of the issues raised that require further investigation, particularly in connection with tutor strategy and tutor style.

We are continuing to investigate research questions related to effective support of exploratory learning in simulation based learning environments (Rosé et al., 2005) as well as continuing to work towards a fully implemented CycleTalk system, with progress reported in (Alevén & Rosé, 2004; Alevén & Rosé, 2005).

ACKNOWLEDGEMENTS

This project is supported by ONR Cognitive and Neural Sciences Division, Grant number N000140410107. We would like to thank Cristen Torrey, who was involved in earlier stages of this work. We would also like to thank our four anonymous tutors.

REFERENCES

- Aleven, V., Koedinger, K. R., & Popescu, O. (2003). A Tutorial Dialogue System to Support Self-Explanation: Evaluation and Open Questions. In U. Hoppe, F. Verdejo & J. Kay (Eds.) *Proceedings of the 11th International Conference on Artificial Intelligence in Education, AI-ED* (pp. 39-46). Amsterdam: IOS Press.
- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R.M. (2003). Seeking and Providing Help in Interactive Learning Environments. *Review of Educational Research*, 73(2), 277-320.
- Aleven, V., & Rosé, C. P. (2004). Towards Easier Creation of Tutorial Dialogue Systems: Integration of Authoring Environments for Tutoring and Dialogue Systems. *Proceedings of the Workshop on "Dialog-based Intelligent Tutoring Systems: State of the art and new research directions"* at the 7th Annual Intelligent Tutoring Systems Conference, Maceio, Brazil, Berlin: Springer Verlag.
- Aleven, V., & Rosé, C. P. (2005). Authoring plug-in tutor agents by demonstration: Rapid tutor development. In C-K. Looi, G. McCalla, B. Bredeweg & J. Breuker (Eds.) *Proceedings 12th International Conference on Artificial Intelligence Education (AIED)* (pp. 735-737). Amsterdam: IOS 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.
- Arguello, J., & Rosé, C. P. (to appear). InfoMagnets: Making Sense of Corpus Data. Companion Proceedings for Human Language Technology conference – North American Chapter of the Association for Computational Linguistics annual meeting (HLT-NAACL) 2006, New York City.
- Baher, J. (1999) Articulate Virtual Labs in Thermodynamics Education: A Multiple Case Study. *Journal of Engineering Education*, October. 429-434.
- Carroll, J., Mack, R., Lewis, C., Grischowsky, N., & Robertson, S. (1985). Exploring a word processor. *Journal of Human-Computer Interaction*, 1, 283-307.
- Charney, D.H., & Reder, L.M. (1986). Designing tutorials for computer users: Effects of the form and spacing of practice on skill learning. *Human Computer Interaction*, 2, 297-317.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science*, 5(2), 121-152.
- D'Souza, A., Rickel, J., B. Herreros, B., & Johnson, W. (2001). An Automated Lab Instructor for Simulated Science Experiments. In J. D. Moore, C. L. Redfield & W. L. Johnson (Eds.) *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future, Proceedings of AI-ED 2001*. (pp 65-76). Amsterdam: IOS Press.
- Evens, M., & Michael, J. (2006). *One-on-One Tutoring by Humans and Machines*. Lawrence Erlbaum and Associates.
- Forbus, K. D., Whalley, P. B., Evrett, J. O., Ureel, L., Brokowski, M., Baher, J., Kuehne, S. E. (1999). CyclePad: An Articulate Virtual Laboratory for Engineering Thermodynamics. *Artificial Intelligence*, 114(1-2), 297-347.
- Graesser, A., Moreno, K. N., & Marineau, J. C. (2003). AutoTutor Improves Deep Learning of Computer Literacy: Is it the Dialog or the Talking Head? In U. Hoppe, F. Verdejo & J. Kay (Eds.) *Proceedings of the 11th International Conference on AI in Education (AI-ED'03)* (pp. 47-54). Amsterdam: IOS Press.
- Graesser, A., VanLehn, K., the TRG, & the NLT. (2002). *Why2 Report: Evaluation of Why/Atlas, Why/AutoTutor, and Accomplished Human Tutors on Learning Gains for Qualitative Physics Problems and Explanations*. LRDC Tech Report, University of Pittsburgh.
- Graesser, A. C., Bowers, C. A., Hacker, D.J., & Person, N. K. (1998). An anatomy of naturalistic tutoring. In K. Hogan & M. Pressley (Eds.) *Scaffolding of Instruction* (pp 145-184), Brookline Books.

- Gweon, G., Arguello, J., Pai, C., Carey, R., Zaiss, Z., & Rosé, C. P. (2005). Towards a Prototyping Tool for Behavior Oriented Authoring of Conversational Interfaces. In J. Burstein & C. Leacock (Eds.) *Proceedings of the ACL Workshop on Educational Applications of NLP*. Alberta, Canada.
- de Jong, T., & van Joolingen, W. R. (1998). Scientific Discovery Learning With Computer Simulations of Conceptual Domains. *Review of Educational Research*, 68(2), 179-201.
- Jordan, P., Rosé, C. P., & VanLehn, K. (2001). Tools for Authoring Tutorial Dialogue Knowledge. In J. D. Moore, C. L. Redfield & W. L. Johnson (Eds.) *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future, Proceedings of AI-ED 2001* (pp. 222-233). Amsterdam: IOS Press.
- Kashihara, A., Kinshuk, Opperman, R., Rashev, R., & Simm H. (2000). A Cognitive Load Reduction Approach to Exploratory Learning and Its Application to an Interactive Simulation-based Learning System. *Journal of Educational Multimedia and Hypermedia*, 9(3), 253-276.
- Kinshuk, & Lin, T. (2003). User Exploration Based Adaptation in Adaptive Learning Systems. *International Journal of Information Systems in Education*, 1(1), 22-31.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1-48.
- Koedinger, K. R., Anderson, J.R., Hadley, W.H., & Mark, M.A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30-43.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The Case for Guided Methods of Instruction. *American Psychologist*, 59(1), 14-19.
- Nückles, M., Wittwer, J., & Renkl, A. (2003). Supporting the computer experts' adaptation to the client's knowledge in asynchronous communication: The assessment tool. In F. Schmalhofer, R. Young & G. Katz (Eds.) *Proceedings of EuroCogSci 03*. The European Cognitive Science Conference (pp. 247-252). Mahwah, NJ: Erlbaum.
- Reimann, P. (1991). Detecting functional relations in computerized discovery environment. *Learning and Instruction*, 1, 45-61.
- Renkl, A. (2002). Learning from worked-out examples: Instructional explanations supplement self-explanations. *Learning & Instruction*, 12, 529-556.
- Rosé, C. P., Jordan, P., Ringenber, M., Siler, S., VanLehn, K., & Weinstein, A. (2001). Interactive Conceptual Tutoring in Atlas-Andes. In J. D. Moore, C. L. Redfield & W. L. Johnson (Eds.) *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future, Proceedings of AI-ED 2001* (pp. 256-266). Amsterdam: IOS Press.
- Rosé, C. P., Bhembé, D., Siler, S., Srivastava, R., & VanLehn, K., (2003b). The Role of Why Questions in Effective Human Tutoring. In U. Hoppe, F. Verdejo & J. Kay (Eds.) *Proceedings of the 11th International Conference on AI in Education, (AI-ED 2003)* (pp 55-62), Amsterdam: IOS Press.
- Rosé, C. P., Torrey, C., Alevén, V., Robinson, A., Wu, C., & Forbus, K. (2004). CycleTalk: Towards a Dialogue Agent that Guides Design with an Articulate Simulator. In J.C. Lester, R.M. Vicari & F. Paraguaçu (Eds.) *Proceedings of the Intelligent Tutoring Systems Conference* (pp. 401-411). Berlin: Springer.
- Rosé, C. P., Alevén, V., Carey, R., Robinson, A., & Wu, C. (2005). A First Evaluation of the Instructional Value of Negotiable Problem Solving Goals on the Exploratory Learning Continuum. In C-K. Looi, G. McCalla, B. Bredeweg & J. Breuker (Eds.) *Proceedings of the 12th International Conference on AI in Education (AI-ED '05)* (pp. 563-570). Amsterdam: IOS Press.
- Rosé, C. P., & Torrey, C. (2005). Interactivity versus Expectation: Eliciting Learning Oriented Behavior with Tutorial Dialogue Systems. In M. F. Costabile & F. Paterno (Eds.) *Tenth International Conference on Human Computer Interaction (INTERACT 2005)*. LNCS 3585. Berlin: Springer.
- Rosé, C. P., & VanLehn, K. (2005). An Evaluation of a Hybrid Language Understanding Approach for Robust Selection of Tutoring Goals. *International Journal of AI in Education*, 15(4), 325-355.
- Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education*, 46(4), 426-445.

- Smith, R., Hipp, D. R., & Biermann, A. (1995). An Architecture for Voice Dialogue Systems Based on Prolog-Style Theorem Proving. *Computational Linguistics*, 21, 3.
- Swaak, J., & de Jong, T. (1996). Measuring intuitive knowledge in science: The development of the WHAT-IF test. *Studies in Educational Evaluation*, 22, 341-362.
- Swaak, J., & de Jong, T. (2001). Learning Vs. System Control in Using Online Support for Simulation-Based Discovery Learning. *Learning Environments Research*, 4, 217-241.
- Tuttle, K., & Wu, Chih. (2001). Intelligent Computer Assisted Instruction in Thermodynamics at the U.S. Naval Academy. *Proceedings of the 15th Annual Workshop on Qualitative Reasoning*, San Antonio, Texas.
- VanLehn, K., Jordan, P., Rosé, C. P., & The Natural Language Tutoring Group. (2002). The Architecture of Why2-Atlas: a coach for qualitative physics essay writing. In S. A. Cerri, G. Gouardères & F. Paraguaçu (Eds.) *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems, ITS 2002* (pp. 158-167). Berlin: Springer Verlag.
- Wang, N., Johnson, L., Mayer, R., Rizzo, P., Shaw, E., & Collins, H. (2005). The Politeness Effect: Pedagogical Agents and Learning Gains. In C-K. Looi, G. McCalla, B. Bredeweg & J. Breuker (Eds.) *Proceedings of the 12th International Conference on AI in Education (AI-ED '05)* (pp. 686-693). Amsterdam: IOS Press.
- White, B., & Frederiksen, J. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence*, 42, 99-157.
- Wu, C. (2002). *Intelligent Computer-Based Engineering Thermodynamics and Cycle Analysis*. New York: Nova Science Publishers.
- Zinn, C., Moore, J. D., & Core, M. G. (2002). A 3-Tier Planning Architecture for Managing Tutorial Dialogue. In S. A. Cerri, G. Gouardères & F. Paraguaçu (Eds.) *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems, ITS 2002* (pp. 574-584). Berlin: Springer Verlag.