Computers & Education 60 (2013) 159–171

Contents lists available at SciVerse ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Does supporting multiple student strategies lead to greater learning and motivation? Investigating a source of complexity in the architecture of intelligent tutoring systems

Maaike Waalkens [a,b,*], Vincent Aleven [a], Niels Taatgen [b]

[a] HCI Institute, Carnegie Mellon University, Pittsburgh, PA, USA
[b] Department of Artificial Intelligence, University of Groningen, The Netherlands

A R T I C L E I N F O

Article history:
Received 22 January 2012
Received in revised form
19 July 2012
Accepted 25 July 2012

Keywords:
Intelligent tutoring systems
Learning strategies
Interactive learning environments
Architectures for educational technology system

A B S T R A C T

Intelligent tutoring systems (ITS) support students in learning a complex problem-solving skill. One feature that makes an ITS architecturally complex, and hard to build, is support for strategy freedom, that is, the ability to let students pursue multiple solution strategies within a given problem. But does greater freedom mean that students learn more robustly? We developed three versions of the same ITS for solving linear algebraic equations that differed only in the amount of freedom given to students. One condition required students to strictly adhere to a standard strategy, the other two allowed minor and major variations, respectively. We conducted a study in two US middle schools with 57 students in grades 7 and 8. Overall, students' algebra skills improved. Contrary to our hypotheses, the amount of freedom offered by the system did not affect students' learning outcomes, nor did if affect their intrinsic motivation. Students tended to use only the standard strategy and its minor variations. Thus, the study suggests that in the early stages of problem-solving practice within a complex domain, an ITS should allow at least a small amount of freedom, validating, albeit to a limited degree, one source of complexity in ITS architectures. To help students develop strategic flexibility, a desirable outcome in many domains, more is needed than letting students chose their own solution strategy within a given problem.

1. Introduction

Intelligent tutoring systems (ITS) are an adaptive learning technology aimed at helping students learn a complex cognitive skill (Koedinger & Corbett, 2006; VanLehn, 2006; Woolf, 2009). Many ITS support *tutored problem solving*, that is, they provide step-by-step guidance as students solve complex problems (VanLehn, 2006). A distinguishing characteristic of ITS is that they provide guidance with respect to the *steps* in a problem, rather than just providing feedback on the final answer. Some researchers argue that guidance at the step level is what makes ITS effective (VanLehn, 2011). ITS are beginning to be widely used in the US. For example, Cognitive Tutors, a type of ITS grounded in cognitive theory and cognitive modeling, are used in mathematics instruction in about 2700 schools (Koedinger & Corbett, 2006). Many evaluation studies have shown that ITS enhance learning, compared to more typical forms of instruction (Beal, Walles, Arroyo, & Woolf, 2007; Graesser, Chipman, Haynes, & Olney, 2005; Koedinger & Aleven, 2007; Koedinger, Anderson, Hadley, & Mark, 1997; Mitrovic, Martin, & Mayo, 2002; Ritter, Kulikowich, Lei, McGuire, & Morgan, 2007; VanLehn, 2011; VanLehn et al., 2005).

While ITS are effective in helping students learn, they have traditionally been quite difficult to build (Murray, 2003). Many factors contribute to the challenge of building an ITS, including the fact that ITS architectures tend to be quite complex. Here we focus on one feature that contributes to ITS complexity, namely, the ability of ITS to support strategy freedom on the part of the student, or, equivalently, the system's ability to recognize multiple student solution strategies within the same problem. For example, in the *Algebra Cognitive Tutor* (e.g., Koedinger et al., 1997; Koedinger & Corbett, 2006), a successful and widely used ITS, students can solve equations using any correct sequence of operations. The tutor able to provide guidance at each step, regardless of which strategy the student decides to follow. This ability is often thought to be an important and attractive characteristic of ITS, even if it is not always called out in theoretical accounts. For example,

* Corresponding author. WA Scholtenstraat 8/8, 9712 KW Groningen, The Netherlands. Tel.: +31 628416582.
E-mail address: maaikewaalkens@hotmail.com (M. Waalkens).

0360-1315/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compedu.2012.07.016

VanLehn's (2006) seminal article about the behavior of tutoring systems assumes that ITS are capable of supporting multiple solution paths within a single problem, but does not call attention to this feature. The ability to support multiple paths is assumed to be necessary in domains where problems have multiple solutions strategies and thought to help students learn more effectively. This ability, however, comes at a considerable cost: systems that support multiple solution paths tend to be considerably more complex than simple tutors that support only a single solution path within any given problem, as discussed further below. But does this particular source of complexity help make ITS more effective? Put differently, does the greater freedom afforded by ITS result in better learning results and motivation on the part of students?

In many domains, one readily finds problems that allow for multiple solution methods. Also, many solution methods allow for small variations, further adding to the variety of problem-solving behavior. For example, in the domain of algebra (the application domain for the current study), even relatively simple equations can be solved in different ways. As another example, even in early math learning, one finds many problems that can be solved in multiple ways. In a fraction addition problem, for example, different common denominators can often be used. In proportional reasoning, many different solution strategies have been identified (e.g., http://en.wikipedia.org/wiki/Proportional_reasoning#Examples) and within each of these strategies, many minor variations are possible. Even a simple solution procedure based on cross-multiplying yields 64 (minor) variations, corresponding to different solution paths. Similarly, in computer programming (another popular domain for ITS development), solutions to even very simple programming problems typically allow for multiple solution strategies, which each have large numbers of many variations. Even in very simple Prolog programs, the number of variations can number in the thousands (e.g., Le & Pinkwart, 2011). Likewise, in domains such as geometry, logic, and physics, problems typically allow for a rather large number of different solutions. In short, problems with multiple solution strategies and variants are tantamount in many domains. To operate effectively in such domains, a tutoring system must, it would appear, be able to recognize all the major and minor variations, lest it flag correct student solutions as incorrect, which may be detrimental for learning and motivation.

To do so, ITS typically represent relevant (domain-specific) problem-solving knowledge and use it to provide tutoring. Different types of ITS use different knowledge representation techniques, but they all use domain-specific knowledge to follow students as they choose their own individual path through the problem space, to evaluate student solution steps or partial solutions, and to provide correctness feedback, hints and error messages. For example, Cognitive Tutors use production rules (e.g., Aleven, 2010; Anderson, Corbett, Koedinger, & Pelletier, 1995; Koedinger & Aleven, 2007; Koedinger & Corbett, 2006) to represent the knowledge and strategies of a competent problem solver in the given domain. In the *Algebra Cognitive Tutor* (Koedinger et al., 1997; Koedinger & Corbett, 2006), different algebraic operations are captured by different production rules; the tutor can generate different solution paths by stringing together different applicable rules, enabling it to recognize a great variety of student strategies. Constraint-based tutors represent the constraints that correct problem solutions must satisfy (e.g., Mitrovic et al., 2002). Typically, many different solutions and solution variants fall within a given set of constraints, and therefore, will be recognized as correct by the tutor. Example-tracing tutors (Aleven, McLaren, Sewall, & Koedinger, 2009; Koedinger, Aleven, Heffernan, McLaren, & Hockenberry, 2004) use behavior graphs, elaborated examples that explicitly enumerate the solution space of problems. As discussed below, different paths in the graphs represent different strategies, enabling the tutor to recognize these different strategies. By flexibly matching student behavior against the graph, they can also recognize many minor variations characterized by different order of steps, small notational variants, and even, different choice of steps (e.g., different common denominators in fraction addition problems). The architecture of each of these systems is complex however, since it must support the particular type of knowledge representation used, and must support its use in service of the various tutoring functions that these systems perform.

On the other hand, several effective ITS support only a single solution path within any given tutor problem, for instance ASSISTments (Feng, Heffernan, & Koedinger, 2009), AnimalWatch (Arroyo, Woolf, & Beal, 2006) and Wayang Outpost (Beal et al., 2007). Such systems tend to have less complex architectures and tend to be easier to build. In theory, single-path tutors can be built by building a tutor interface for the given problem type, and associating tutoring support with specific interface elements. The support associated with a given interface element may include correct and incorrect answers, as well as hint and error messages. One reason that this simple approach works is that if there is only one solution path, the answer to one-step in a given problem does not vary depending how a prior step in that same problem has been answered.

ITS Researchers and developers have so far assumed that problem-solving activities amenable to multiple solution strategies, freedom and supporting multiple strategies is beneficial for learning results and motivation (Mitrovic, Mayo, Suraweera, & Martin, 2001; VanLehn, 2006). Mastery, understanding, and flexible use of multiple strategies are often seen as an important ingredient of "adaptive expertise" (Hatano & Inagaki, 1986). In mathematics education (the current study focuses on algebraic problem solving), the importance of focusing students on multiple strategies is highlighted in many sources, including the NCTM standards (NCTM, 2000, p. 64) and a book reviewing research advances (Kilpatrick, Swafford, & Findell, 2001). A study by Ainsworth, Wood, and O'Malley (1998) found that requiring children to produce multiple answers to math problems was effective for a low-performing subset of students. Multiple researchers have pointed out that an instructional approach in which students use and reflect on multiple strategies is prevalent in countries that typically do very well in international comparisons of math competence, such as China, and is less prevalent in countries that are not as highly ranked (e.g., Ma, 1999; Stigler & Hiebert, 1999).

Offering freedom with respect to strategies may offer both cognitive and motivational benefits. With respect to cognitive benefits, when an ITS offers strategy freedom, students may try out different strategies, experience when each strategy is effective, and develop intuitions as to when each strategy is most effectively, in short might (ideally) help students develop strategic flexibility, as discussed above, an important element of expertise in any domain. Supporting freedom may also be motivating; strategy freedom gives students options to choose from and different studies show that "choice" is important for a higher intrinsic motivation and learning (Cordova & Lepper, 1996; Patall, Cooper, & Wynn, 2010). Conversely, not supporting strategies may lead to unnecessary confusion, for instance, when the system does not accept an alternative strategy that a student happens to know about. It may also be uninspiring or demotivating (e.g., a student may be disappointed when the system rejects a particularly clever but unusual choice of solution strategy), or quite possibly frustrating, with a likely detrimental effect on students' learning outcomes.

In spite of the importance of the issue, and the strong arguments that can be made in favor of supporting multiple strategies, we are not aware of any empirical studies in the literature on ITS that ask whether supporting multiple strategies does indeed pay off in terms of more robust learning on the part of students. The current article describes an experiment to address this issue. Specifically, the goal of our study is to investigate to what extent the amount of freedom students experience in a tutor leads to more robust learning and greater intrinsic motivation.

We chose to investigate this research question in the domain of early algebra learning, one of the many domains characterized by multiple solution paths. Algebra is a suitable task domain for research into freedom and restrictions, because a linear equation (e.g., $2(x + 1) + 1 = 5$) can often be solved in multiple ways. On the other hand, there is a standard strategy (discussed below) that is widely used in US schools. This strategy can solve many linear equations in an optimal way (i.e., with a minimum number of solution steps), but occasionally yields a solution path that is slightly longer than strictly necessary. We developed three versions of an ITS for solving linear equations. These versions differed only in the amount of freedom offered within each problem (or equivalently, the range of solution paths that the tutor recognized as valid solutions). One tutor version accepted only the standard solution strategy, one accepted the standard strategy with minor variations, and a third version accepted all reasonable strategies, as long as each solution step was closer to the solution.

Star (2005) suggests that there is a possible trade off in initial stages of learning between the goal of flexible use of multiple strategies and the goal of mastery of a standard algorithm. Star and Rittle-Johnson (2008) showed that prompting students to solve the same equation in different ways provides better results on items measuring students' strategic flexibility. However, students' procedural knowledge improved less. Therefore, we expect that more restricted tutors can help students to learn a well-defined, optimal problem-solving strategy and freer tutors can be more helpful for deeper understanding.

2. Equation-solving strategies

To address our research questions, we developed three versions of an ITS for solving linear equations, each with exactly the same sequence of equations for students to solve. The ITS contains a total of 44 equations divided into 5 categories, shown in Table 1. These types of equations are similar to those found in several textbooks. The equations are sequenced in order of increased complexity, as indicated in Table 1.

The three versions of the ITS differ only in the level of freedom they allow and support. The versions are: (a) *strict standard strategy*, (b) *flexible standard strategy* or (c) *multi strategy*. In the two *standard strategy* conditions, all equations had to be solved with a standard strategy that is widely used in American middle-school mathematics textbooks (Benson et al., 1991; Holt, Rinehart, & Winston, 2007). This strategy can solve almost all linear equations. The strategy is as follows: First, use the distributive law to expand any term in parentheses. Second, combine constant terms and variable terms on each side of the equation. Third, move variable terms to one side of the equation and constant terms to the other side. And finally, divide both sides by the coefficient of the variable term.

In the *strict standard strategy*, further restrictions were imposed. First, students must show intermediate steps while solving equations, in which they indicate the particular transformation being used. For example, in the equation $(2x) + 1 = 5$, the only acceptable next step is $2x + 1 - 1 = 5 - 1$. The ITS does not allow the student to go from $2x + 1 = 5$ to $2x = 4$ without showing the intermediate step. In the American schooling system, students just learning to solve equations are often required by their teachers to write out these intermediate steps. A further constraint imposed in the *strict standard strategy* concerns so-called "double move operations." When an equation has a constant and a variable term on both sides of the equation (e.g., $2x + 2 = x + 4$), two move operations are needed in order to collect the variable terms on one side of the equation and the constant terms on the other side, leading to four minor variations of the strategy (see Table 2). In the *strict standard strategy,* only one of these variations is allowed: the student must start by moving the smallest variable term, which is most common in mathematics textbooks (Benson et al., 1991; Holt et al., 2007). An example is shown in the left column of Table 2.

In the *flexible standard strategy* condition, on the other hand, students use the standard strategy without these restrictions. All minor variations are allowed, including the four options for double move steps shown in Table 2. Further, students are free to either skip or do the intermediate steps described above.

Students have the most freedom in the *multi-strategy condition*; students can solve the equations with any strategy that progresses toward the goal of solving the equation. For instance, in the equation $(2)(x + 1) = 4$ students are allowed to divide both sides by 2, instead of using the distributive law to expand the term in parentheses, which is required in the two stricter conditions. The only restriction is that an operation is allowed only if it brings the student closer to the solution. In the equation $(2x) + 1 = 5$ it mathematically correct to add 100 to both sides of the equation, but this operation does not represent progress toward a solution, so this step is not allowed. A second restriction is that a divide step is allowed only when it does not generate fractions. Given the time available for the study, we limited the tutor curriculum to equations with whole numbers. Despite these two restrictions, the *multi-strategy* condition covers all reasonable strategies. Importantly, it allows the use of more optimal strategies than the standard strategy (with which equations can be solved faster/with fewer problem-solving steps). Examples of all allowed strategies (including ones more optimal than the standard strategy) are shown in Tables 3 and 4.

Table 1
Equations in the ITS.

Day	Equations	Example	Number
Day one	One-step	$x + 2 = 5$	6
	Two-steps	$3x + 1 = 7$	6
	Parentheses	$2(x + 1) = 4$	3
Day two	Parentheses	$2(x + 1) = 4$	4
	Parentheses, more difficult	$2(x + 1) + 1 = 5$	8
	Multiple steps	$3x + 1 = 2x + 3$	8
Day three	Multiple steps	$3x + 1 = 2x + 3$	5
	Parentheses, more difficult	$2(x + 1) + 1 = 5$	4
Total			44 problems

Table 2
Minor variations of the *standard strategy*. All variations are allowed in the *flexible standard strategy* condition and the *multi-strategy* condition. Only the first (shown in the leftmost column) is allowed in the *strict standard strategy*. In the *strict standard strategy* intermediate steps are required, although they are not shown in this example.

Standard start with lowest variable term	Alternative 1 start with highest variable term	Alternative 2 start with lowest constant term	Alternative 3 start with highest constant term
$3x + 4 = x + 6$	$3x + 4 = x + 6$	$3x + 4 = x + 6$	$3x + 4 = x + 6$
$2x + 4 = 6$	$4 = -2x + 6$	$3x = x + 2$	$3x - 2 = x$
$2x = 2$	$-2 = -2x$	$2x = 2$	$-2 = -2x$
$x = 1$	$1 = x$	$x = 1$	$1 = x$

3. System design

As is typical of ITSs (e.g., VanLehn, 2006), the equation-solving tutor developed for our study provides step-by-step guidance during students' problem-solving activities, including correctness feedback, next-step hints, and error feedback messages related to common student errors. When the student enters a step, the tutor provides correctness feedback. Correct answers turn green, incorrect answers turn red. Students must explain their step by selecting from a menu, as illustrated in Fig. 1. Menu-based self-explanations such as those supported by the equation-solving tutor have been shown to be effective in prior research on ITS (e.g., Aleven & Koedinger, 2002). At any time, the tutor provides context-specific hints at the students' request. Usually, multiple levels of hints are available for any given step in a tutor problem, more general hints first, followed by increasingly specific hint levels. The last hint level gives the step to the student. The hints are the same in all three versions and always focus on the standard strategy, except when, in the *multi-strategy* condition, a student decides to deviate from the standard strategy; when they do, the hints focus on the chosen strategy.

For some common errors, the tutor displays a specific error feedback message, with information about the error and how to correct it. There are specific error feedback messages for steps deemed "wrong" in the two stricter conditions that are allowed in the free-est condition. These messages explain that the step is not mathematically wrong, but that the standard strategy should be followed, as illustrated in Fig. 2. (We will use the term "unallowed valid variations" to denote such errors.) Finally, there is an introduction screen with instructions at the start of each problem set. This screen shows an example of how to solve the kind of equations included in that set, with explanations.

All conditions were implemented as example-tracing tutors (Aleven, McLaren, Sewall, & Koedinger, 2009), a type of ITS that is relatively easy to implement, given that efficient and mature authoring tools exist, as long as the number of solution paths is limited. Example-tracing tutor support the same range of tutoring behavior as other types of ITS, including Cognitive Tutors (Aleven, McLaren, Sewall, & Koedinger, 2009; VanLehn, 2006). Example-tracing tutors represent the solution space for a given problem as a *behavior graph*. The behavior graph maps out the solution paths for the given problem, as well as some common errors. The tutor evaluates students' problem-solving steps by comparing them to steps in the behavior graph. Essentially, it checks, in a flexible way, whether the students' solution steps correspond to any path through the graph and gives correctness feedback accordingly. It also uses the graph to select hints and error feedback messages to present to the students. A brief overview of the example-tracing algorithm can be found in Aleven, McLaren, Sewall, and Koedinger (2009). Example-tracing tutors can be created without actual coding. First, a tutor interface is created through drag-and-drop techniques. Second, behavior graphs are created "by demonstration", that is, simply by entering solution steps in the tutor interface. The steps are recorded automatically in a behavior graph. Building example-tracing tutors with CTAT was shown in a range of projects to be 4–8 times faster than building a Cognitive Tutor (Aleven, McLaren, Sewall, & Koedinger, 2009).

A measure of the complexity of the different tutor versions is the number of paths in each behavior graph, shown in Fig. 3. (In all example-tracing tutors, each tutor problem has its own graph; problems of the same type have isomorphic graphs.) The *strict standard strategy* condition had only a single path per graph (Fig. 3, left). The links branching off the main path represent errors and enable the tutor to recognize these specific errors and provide feedback on them. The *flexible standard strategy* condition had a slightly larger number of paths in the behavior graph (Fig. 3, center). Finally, the behavior graphs for the most complex problems in the *multi-strategy* condition contain 96 paths[1] (Fig. 3, right).

Incidentally, although the example-tracing technology was fully adequate for implementing the tutoring behavior for all three tutoring conditions (i.e., step-by-step guidance during problem solving), we felt that it was an optimal choice of technology only for the *flexible standard strategy* condition, our middle condition, and not an ideal choice for the other two conditions. In particular, in implementing the most complex tutor version (*multi-strategy* condition), it was a substantial and painstaking effort to capture (in a behavior graph) the large number of different strategies that the tutor needs to support. This tutor is definitely on the far end of complexity that the example-tracing technology can conveniently handle. For building real-world equation-solving tutors, one might prefer to use rule-based based methods (as used in e.g., the *Algebra Cognitive Tutor* (Koedinger et al., 1997; Koedinger & Corbett, 2006)). However, all things considered, we do believe that using the example-tracing technology for all conditions in our experiment was a good choice. First, in experiments comparing different tutor versions, it is often better to keep the technology constant across conditions, so as to minimize differences between conditions other than those whose effect one wants to evaluate. Second, the authoring tools used to create example-tracing tutors (CTAT, Aleven, McLaren, Sewall, & Koedinger, 2009) support a range of other functionality besides the actual tutoring

[1] In the *multi-strategy* condition, the behavior graph for *equations with parentheses more difficult* contains 96 solution paths. There are 7 different strategies to use (see Table 4). Each strategy has multiple paths because 3 of the 4 operations (except combine like terms) can be done in two different ways: with or without showing intermediate steps. The standard strategy (Table 4, left) has four operations and thus 2*1(combine like terms)*2*2 = 8 different paths. Alternative 1 (Table 4, second left) has three operations and 2*2*2 = 8 different paths. Alternatives 2 though 6 have four operations and 2*2*2*2 = 16 different paths each. So there are in total 2*8 + 5*16 = 96 different paths (for the graphic representation, see Fig. 3, right).

Table 3
All possible strategies in the *multi-strategy* condition for *equations with parentheses*. In addition to the standard strategy (shown in the leftmost column), two alternative strategies are allowed, one of which is more efficient than the standard strategy (alternative 1). Intermediate steps are optional in the *multi-strategy* condition.

Standard strategy	Alternative 1 more optimal	Alternative 2
$2(x+1)=4$	$2(x+1)=4$	$2(x+1)=4$
$2x+2=4$	$x+1=2$	$2x+2=4$
$2x=2$	$x=1$	$x+1=2$
$x=1$		$x=1$

behavior that is very important in classroom experiments, such as web deployment, learning management facilities, and detailed logging of student–tutor interactions.

4. Experiment

We conducted a study to evaluate the effect of these three levels of freedom on student learning and motivation. The study took place during the summer holidays in two schools in a school district near Pittsburgh, with assistance from teachers at the school.

4.1. Participants

73 Students participated in the study and we obtained valid data for 57 students, 28 of whom were entering grade 7 and 29 of whom were about to enter the 8th grade. Not included in this group of 57 are students who were not present for all tutoring, as well as the students who completed less than half of the problems in the ITS and students who attempted fewer than 1/4th of the problems on the post-test, considered to be a lack of effort that rendered their post-test results useless.

Participation was voluntary. The students were recruited by means of an e-mail message sent by the principal of the school to the parents.

4.2. Procedure

All students participated three consecutive days, 2 h a day. Students were randomly assigned to one of the three conditions described above. They worked on the tutor during all three days, with a paper pre-test at the beginning of day 1 and a paper post-test and motivation questionnaire at the end of day 3. Each day, the students' worked on different problem sets (see Table 1). If a student completed the daily set of equations before the end of the session, they worked on another problem set unrelated to algebra. All tutors were deployed on our tutoring website, *Mathtutor* (http://mathtutor.web.cmu.edu, see Aleven, McLaren, & Sewall, 2009).

4.3. Measures

The paper-based pre-test and post-test were designed to measure various aspects of students' algebra learning (see Table 5). The pre-test had 20 questions and was a subset of the post-test, which contained 42 questions. We made two nearly isomorphic versions of the 20 items that were shared between pre-test and post-test, and used two test forms that were used in a counterbalanced manner across pre-test and post-test.

We wanted to keep the pre-test small for several reasons: First, we had limited time to conduct the study. Second, we did not want to frustrate students by assigning a test that was too difficult. Third, we expected the student performance for the more difficult items to be low at pre-test, so assigning these items might not have yielded useful information. Both tests assessed procedural knowledge, conceptual knowledge and flexibility in problem solving. The procedural items were divided in two categories, familiar equations (equations like those encountered in the ITS) and transfer equations (with a feature not practiced in the ITS). Conceptual items tested the understanding of the principles that govern algebra and of their interrelations, for example determining if two equations are equivalent. Flexibility items measure students' facility with, and knowledge about different (and more efficient) strategies. For example, some of these items tested if students were able to recognize the validity and correctness of strategies other than the standard strategy. There were also items where students had to solve an equation twice, with two different strategies. To ensure face validity, many test items were similar to assignments in middle-school mathematics text books (Benson et al., 1991; Holt et al., 2007). Also, experienced middle-school mathematics participated in the test construction processes and helped make sure that the test was appropriate and at an appropriate level of difficulty.

Table 4
All possible strategies in the multi-strategy condition for *equations with parentheses, more difficult*. In addition to the standard strategy (shown in the leftmost column), a range of other strategies is allowed.

Standard strategy	Alternative 1 more optimal	Alternative 2	Alternative 3	Alternative 4	Alternative 5	Alternative 6
$2(x+1)+1=5$	$2(x+1)+1=5$	$2(x+1)+1=5$	$2(x+1)+1=5$	$2(x+1)+1=5$	$2(x+1)+1=5$	$2(x+1)+1=5$
$2x+2+1=5$	$2(x+1)=4$	$2(x+1)=4$	$2(x+1)=4$	$2x+2+1=5$	$2x+2+1=5$	$2x+2+1=5$
$2x+3=5$	$x+1=2$	$2x+2=4$	$2x+2=4$	$2x+2=4$	$2x+2=4$	$2x+1=3$
$2x=2$	$x=1$	$2x=2$	$x+1=2$	$x+1=2$	$2x=2$	$2x=2$
$x=1$		$x=1$	$x=1$	$x=1$	$x=1$	$x=1$

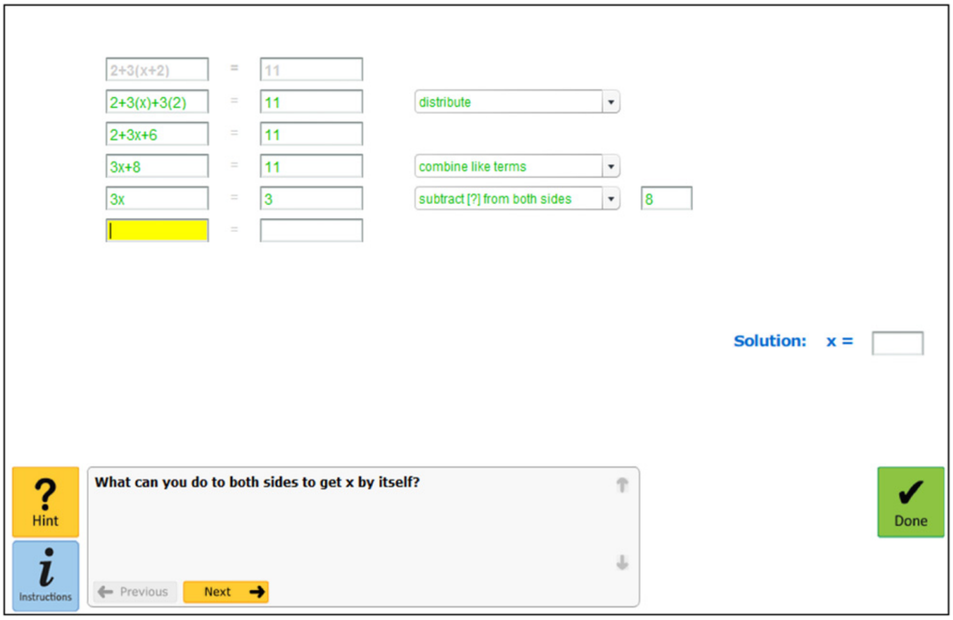

Fig. 1. Screenshot of the ITS. Students solve equations with step-by-step guidance from the tutor, such as correctness feedback (correct steps colored green) and hints. Also, after each operation students explain what they have done, using short menus in the tutor interface. In the last step the student used the freedom to not show intermediate steps. The tutor hints (an example of which is shown in the panel at the bottom, next to the "Hint" button) are context specific and available at the students' request. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In order to measure aspects of students' motivation, we used the Intrinsic Motivation Inventory (IMI), an existing and validated motivation questionnaire. The IMI is based on self-determination theory, which represents a broad framework for the study of human motivation and personality (e.g., Ryan, 1982; Ryan, Mims, & Koestner, 1983). The IMI was developed to assess participants' subjective experience related to a target activity. The (standard) IMI questions can be modified slightly to fit specific activities without effecting its reliability or validity. For example, we changed the standard question "I tried very hard on *this activity*" to "I tried very hard on *this math workshop*".

There were a total of 26 questions, all involving a 7 point Likert-scale. Questions were divided in four categories, (a) Effort/Importance, (b) Interest/Enjoyment, (c) Perceived Competence, and (d) Value/Usefulness. We added three questions about tutor-supported flexible strategy use and the perceived freedom in the tutor.

Finally, we collected and analyzed tutor log data. These logs record all student–tutor interactions in substantial detail. We analyzed these data using DataShop, an open repository for interaction data from educational technology (Koedinger, Cunningham, Skogsholm, & Leber, 2008). DataShop is seamlessly integrated with CTAT, the authoring tools we used to construct the equation-solving tutor. We used the tutor log data to investigate (a) evidence of within-tutor learning, (b) students' strategy use, and (c) the nature of students' strategy-related errors.

4.4. Hypotheses

Previous research lead to the hypothesis that more restricted tutors can help students to learn a well-defined, optimal problem-solving strategy and freer tutors can be more helpful for deeper knowledge and transfer learning. Star (2005) mentioned possible trade offs in initial

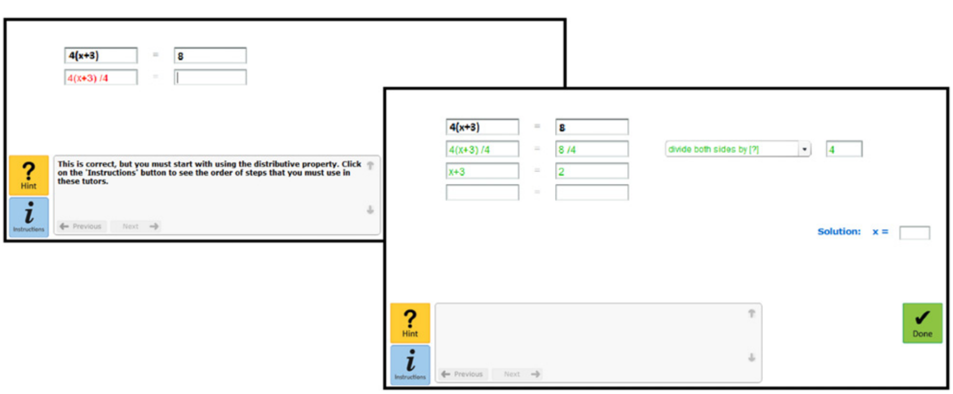

Fig. 2. An action that is flagged as "wrong" in the *strict standard strategy* condition (left; the step is colored red) but not in the *multi-strategy* condition (right). The tutor presents an error-specific feedback message in response to such errors (shown on the left, in the hint panel at the bottom). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

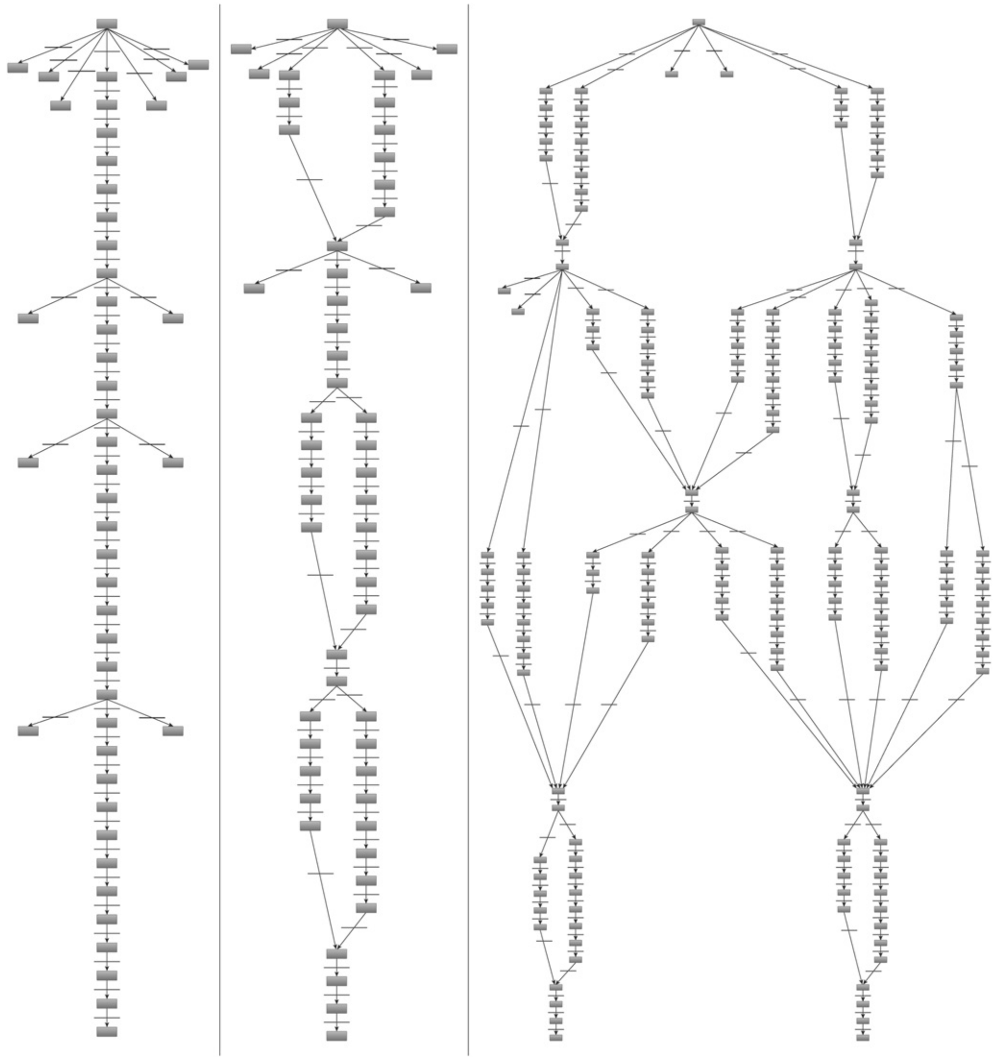

Fig. 3. Three behavior graphs for the same problem (for more difficult equations with parentheses, e.g., $2(x+1)+1=5$). From left to right: *strict standard strategy*, *flexible standard strategy* & *multi strategy*.

stages of learning between the goal of flexible use of multiple strategies and the goal of mastery of a standard algorithm and the three versions differ in their ability to practice one or the other skill. Therefore, we expected that students in the two stricter conditions (*strict standard strategy* and *flexible standard strategy*) would score better on solving familiar equations (i.e., equations of the type practiced in the ITS). These students are forced to use a standard strategy and therefore they will practice this strategy extensively. On items for which deeper understanding is needed (flexibility and transfer items) we expected better results for students in the free-est condition (*multi strategy*) because students can try and practice different strategies, which may support flexible thinking and are possibly better for deeper understanding. Moreover, we anticipated that students in the strict(er) conditions have a lower intrinsic motivation because these students are forced to follow a defined sequence of steps, and may get bored or frustrated because alternative thinking is not accepted.

Table 5
Test items. Transfer equations, (added) motivation questions and a large part of the flexibility items were post-test only.

Test item	Example	Pre-test	Post-test
Procedural knowledge Familiar equations	$4x+3=11$	7	7
Procedural knowledge Transfer equations	$0.2x+0.7=1.1$	–	4
Conceptual knowledge	$3-4x$ is equivalent to $-4x+3$ true/false	8	8
Flexibility items	$4x+2=8$	5	23
	Divide both sides by 2 good move/not a good move		
	Subtract 2 from both sides good move/not a good move		
	Add 2 to both sides good move/not a good move		
Intrinsic motivation questions	I tried very hard on this math workshop.	–	23
	I enjoyed this math workshop very much.		
Added motivation questions	The tutor let me solve the problem the way I wanted.	–	3

Table 6
Estimated marginal means (proportion of correct items) and standard deviations (in parentheses) of all items that were repeated between pre-test and post-test.

Test	Items	Strict standard (18 students)	Flexible standard (16 students)	Multi (23 students)
Pre-test	Procedural familiar	0.52 (0.31)	0.57 (0.30)	0.61 (0.27)
	Conceptual	0.47 (0.28)	0.58 (0.27)	0.64 (0.23)
	Flexibility	0.34 (0.27)	0.46 (0.29)	0.45 (0.24)
Post-test	Procedural familiar	0.64 (0.26)	0.63 (0.24)	0.66 (0.41)
	Conceptual	0.60 (0.20)	0.55 (0.22)	0.66 (0.20)
	Flexibility	0.33 (0.24)	0.48 (0.22)	0.43 (0.25)

5. Results

Two graders graded every pre- and post-test. All items were graded as either correct or incorrect. Tables 6 and 7 show the means and standard deviations. Cronbach's α was 0.76 for the pre-test and 0.87 for the post-test, indicating acceptable to good reliability for the algebra test instrument.

To analyze the learning gains from pre-test to post-test, as well as differences between conditions in these learning gains, we used repeated measures ANOVAs (with factors "time" and "condition"). These analyses included only the items that were repeated between pre-test and post-test (i.e., on the 20 questions that were essentially the same in the pre- and post-test). Procedural learning gain is measured with familiar equations only (transfer equations were post-test only). On familiar equations, there was a significant main effect for test $F(1,55) = 6.235$, $p = 0.016$. Students show significant improvement from pre- to post-test. On the conceptual items, students did not improve significantly from pre- to post-test. Likewise, there was no improvement on the subset of flexibility items that were included in both the pre-test and the post-test. (Most flexibility items were asked only during the post-test, so learning gain on these types of items was not measured.) There were no reliable differences between the conditions in leaning gain on familiar equations and on conceptual items.

To analyze differences between the conditions at the post-test, we used one-way ANOVAs. (This analysis includes the 22 items that appeared only on the post-test.) With respect to algebra learning, there were no significant differences between the conditions on transfer and (post-test only) flexibility items. With respect to motivation, there were no significant differences between the conditions on the motivation questionnaire and the questions about perceived freedom. Table 8 shows the means and standard deviations of the motivation questionnaire.

In addition to analyzing the data on learning and motivational outcomes, we analyzed the tutor log data. As mentioned, these data include all student actions and tutor responses. First, we studied whether the log data provide evidence of within-tutor learning, to corroborate the learning gains established by the pre-test/post-test analysis. As is customary in ITS research, we analyzed within-tutor learning by analyzing the learning curves (Koedinger et al., 2008, 2011; Stamper et al., 2011). The curves "describe performance at the start of training [and] the rate at which learning occurs …" (Koedinger & Mathan, 2004). They visualize how student performance changes over successive practice opportunities within the tutor, subdivided by the knowledge components that make up the overall skill (see Fig. 4). The learning curves are based on a knowledge component model that identifies the detailed components that make up equation-solving skill. This model was identified based on cognitive task analysis, as is customary in ITS research and development. To illustrate the fine grain size of these knowledge components, examples of knowledge components are "add a constant term to both sides" or "subtract a variable term from both sides". The empirical learning curve shown in Fig. 4 (solid line) summarizes the log data: it shows, for each opportunity, the percentage of students who did *not succeed on their first attempt* (averaged over all knowledge components). To better show the trend in the empirical learning curve, a predicted learning curve was obtained by fitting an Additive Factor Model[2] (Cen, Koedinger, & Junker, 2007) to the learning curve data. It is smoother than the empirical learning curve because much of the noise is filtered out by applying the model.

For all skills together the learning curve decreases, as shown in Fig. 4. We used logistic regression (a linear mixed effect model) to analyze the predicted learning curve. The slope of this curve is significantly different than 0 ($p < 0.001$). Thus, the learning curves indicate that students' skill of solving equations improves while working with the tutor, corroborating the learning gains observed from pre- to post-test.

We also analyzed the log data to investigate the range of strategies used by the students during their work with the tutor. Specifically, we looked at how frequently students used minor variations and major variations of the *strict standard strategy,* in the conditions in which these variations are allowed. As mentioned, we identified two different types of *minor* variations with respect to the *strict standard strategy.* First, in the *strict standard strategy* condition, the students are required to write out intermediate steps, as defined above. In the other two conditions, these intermediate steps may be omitted. In 83% of the steps where students had the freedom to either show or omit intermediate steps, they did show them. In 17% of the steps, therefore, students followed minor variations by omitting the intermediate steps. A second type of minor variation pertains to the order in which double move steps are executed. In this study, double move steps occur only in multi-step equations. In the *strict standard strategy* conditions, students must move the smallest variable term first, in the other condition, students have three more options. On 62% of the multiple step equations (e.g., $3x + 1 = 2x + 3$) the students started by moving the smallest variable term, as is required in the *strict standard strategy*. On 38% percent, therefore, they followed minor variations. Thus, small variations within the standard strategy were used with some regularity, although in a minority of the cases in which they *could* be used.

We also looked at the frequency of *major* variations of the strict standard strategy, or rather, strategies other than the standard strategy. Such strategies were accepted only in the *multi-strategy* condition. In this study, different (and more efficient) strategies are only possible in

[2] With the Additive Factor Model algorithm (AFM), a logistic regression is performed over the "error-rate" learning curve data. The AFM uses a set of customized Item-Response models to predict how a student will perform for each skill on each learning opportunity. Here, a standard regression bounded between 0 and 1, attempts to find the best-fit curve for error-rate data, which also ranges between 0 and 1. More information about the model can be found in Cen et al. (2007) and on the *DataShop* website (http://pslcdatashop.org).

Table 7
Estimated marginal means (proportion of correct items) and standard deviations (in parentheses) of all items that were only on the post-test.

Items	Strict standard (18 students)	Flexible standard (16 students)	Multi (23 students)
Procedural transfer	0.15 (0.21)	0.28 (0.31)	0.30 (0.32)
Flexibility	0.53 (0.15)	0.60 (0.20)	0.60 (0.16)

the two types of equations with parentheses. In only 4% of the equations with parentheses (e.g., $2(x+1)=4$) did students use a strategy that is different from (and more efficient than) the standard strategy. In 9% of the difficult equations with parentheses (e.g., $2(x+1)+1=5$), students used strategies other than the standard strategy. However, the more efficient strategy was not used. In sum, major variations were rarely used, even when the (*multi-strategy* condition) tutor allowed them.

Finally, we looked at how frequently the tutor disallowed mathematically valid strategy variations in the two restricted conditions (i.e., marked, as wrong, student actions that would be allowed in the free-est condition – this analysis could be viewed as the flip side of the previous analysis; rather than looking at allowed deviations from the *strict standard strategy* in conditions in which these deviations are allowed, we look at instances where reasonable deviations are disallowed). These types of "errors" occurred mostly because students tried minor variations of the standard strategy, especially on equations involving double move steps. Disallowed valid double move operations account for 9.2% of all errors in equations with double move operations. 1.1% of all errors were for not showing intermediate steps, another form of valid minor variation that was not allowed in the strictest condition. Disallowed valid *major* strategy variations were very rare (i.e., student use of major alternative strategies in the conditions in which they were not allowed). They account for only 1.4% of errors in equations for which multiple strategies were possible. Overall, the tutor disallowed more mathematically valid strategy variations in the *strict standard strategy* ($m = 6.67$, sd $= 4.03$) than in the *flexible standard strategy* ($m = 0.75$, sd $= 1.18$), $t(20.242) = 5.948$, $p < 0.001$.

6. Discussion

Students in all conditions learned from the equation-solving tutor, as evidenced by pre/post learning gains on the types of equations that students solved with the tutor, as well as the declining errors rates seen in the learning curves extracted from the tutor log data. To our surprise, however, offering students more freedom during problem-solving practice had no discernible effect on their learning gains, intrinsic motivation, or even their perceived level of strategy freedom. Our analysis of the tutor log data indicates that students use the freedom that is offered by the two freer versions only to a limited degree. Small variations within the standard strategy were used regularly, but major variations (i.e., strategies other than the standard strategy) were rarely used, even on equations that are somewhat inviting of such strategies, and even in a tutor condition that allowed such variations.

The lack of strategy variations across conditions explains the lack of differences in the main outcome variables (learning, motivation, perceived strategy freedom). Given that there was virtually no difference between the students' problem-solving strategies in the different conditions, one would not expect to see differences between the conditions in these variables. The finding that students in our study did not explore the different strategies on their own is consistent with those of several math education studies in algebraic equation solving (Alibali, 1999; Star & Rittle-Johnson, 2008).

It is likely that the students' tendency to stick to the standard strategy, even when they have the freedom to use other strategies, reflects their prior algebra instruction: American middle-school math textbooks, and quite possibly American middle-school math teachers, emphasize this strategy. The restricted tutor versions therefore may have matched their prior instruction particularly well (as was our intention as we designed the tutor, as described above). However, we find it hard *not* to view the results as implying (or at least suggesting) that students' natural tendency is to not explore alterative strategies on their own, but to stick with known methods. It is possible that students do not realize that it is useful to search for multiple ways of solving a problem (and find the best way), although we should point out that in this study, there was no particular incentive to do so. It was possible to solve all equations in this study with the standard strategy. Trying alternatives would probably make the task harder. It is quite possible that if the participants, as part of their prior instruction, had been used to greater strategy freedom, or had been exposed more to the fact that linear equations can often be solved in multiple ways, the result of the experiment may have been different. As it stands, however, the results are consistent with prior studies that found that a focus on mathematical understanding as being able to see connections between multiple ways of doing things is not prevalent in much mathematics teaching. Some of this work is discussed in the introduction section (e.g., Ma, 1999; Stigler & Hiebert, 1999).

Let us return to the question that motivated the research: Does the complexity needed in ITS architectures in order to support multi-path problems pay off? Our short answer is: Yes, it does, although the evidence was not as overwhelming as we thought it would be. The experiment showed distinct advantages for supporting minor strategy variations, and the architectural complexity needed to support minor variations is no different from that needed to support major alternative strategies (e.g., both need behavior graphs with multiple branches; or production rules, or constraints). Support for this conclusion comes primarily from the tutor log data.

Although there were no differences in learning gains or motivational outcomes between the conditions, the tutor log data did show advantages for the more flexible conditions, in particular, the *flexible standard strategy* tutor. The tutor log data indicate that students *do* use minor variations within the standard strategy. Students omitted intermediate steps in 17% of the cases and used an alternative double move step in 38%. Such minor variations would be marked wrong in the *strict standard strategy* version of the tutor, but would be accepted in the

Table 8
Estimated marginal means and standard deviations (in parentheses). A score close to 1 indicates low motivation and a score close to 7 high motivation.

Motivation questionnaire	Strict standard	Flexible standard	Multi
Effort/importance	5.48 (1.20)	5.36 (1.42)	5.53 (1.22)
Interest/enjoyment	4.38 (1.27)	3.71 (1.13)	3.78 (1.30)
Perceived competence	4.97 (0.92)	4.86 (1.14)	5.35 (1.55)
Value/usefulness	5.29 (1.33)	4.42 (1.74)	4.92 (1.45)

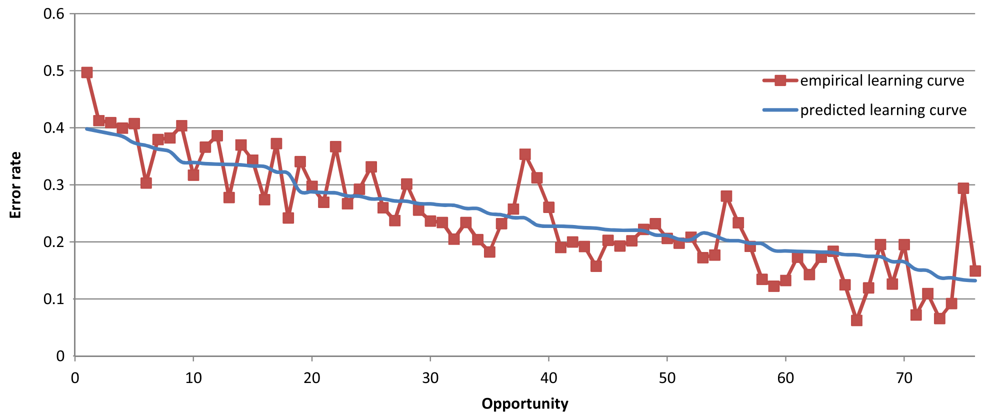

Fig. 4. The predicted error decreases over the learning opportunities.

other tutor versions. Disallowed valid strategy variations were more frequent in the *strict standard strategy* than in the *flexible standard strategy* (and the difference was statistically significant). Although the absolute number of valid strategy variations flagged as errors in the stricter tutor conditions was relatively small, most likely this number underestimates students' natural tendencies to use minor or major variations of the standard strategy. It reflects not only these tendencies but also the students' inclination to follow the tutor's advice. (E.g., after being exhorted by the tutor to include intermediate steps, students are more likely to do so in subsequent steps.) With respect to students' use of major strategy variations, the different strategies allowed in the free-est of the three conditions (i.e., the *multi-strategy* condition) were hardly used.

It may be clear from this analysis that in early algebra learning, a strict single-path tutor is too limiting, as expected, but, much more surprisingly, a fully flexible multi-path tutor appears to be more than is needed, at least to accommodate students' natural strategic tendencies (but see the discussion below). The downside of a strict single-path tutor is that many errors are flagged that, strictly speaking, are not errors (i.e., they are mathematically valid strategy variations). It is important to avoid such spurious errors. Novices face a challenging task in learning equation solving, or any other complex cognitive skill. This task should not be made even harder by unnecessarily flagging, as incorrect, inconsequential deviations from a standard strategy. Further, an ITS that is perceived to be overly strict may not be accepted by teachers. Therefore, there are good reasons to prefer *flexible standard strategy* over the *strict standard strategy* condition: students do use the freedom that is built in, they learn important algebra skills, and the development time is reasonable. Put differently, even for beginning algebra learners, a tutor that supports minor strategy variations is preferred over one that strictly enforces a standard strategy. The study therefore provides support for this source of architectural complexity in ITS, although not as overwhelmingly as hypothesized.

It is harder to find support in our data for a fully flexible tutor, compared to one that allows a standard strategy and minor variations. Students in the *multi-strategy* condition did not use multiple strategies, even though the tutor allowed them to do so. Apparently, merely offering the flexibility to use multiple strategies does not mean that students will take advantage of this flexibility. We would however strongly caution against an implication that tutors with a limited amount of freedom are *always* sufficient. A tutor that supports only a standard strategy and its minor variations may be helpful early on during algebra instruction, when the focus is on acquiring basic skills and understanding. It is very likely however that systems that support multiple strategies are necessary in order for students to develop genuinely robust and flexible problem-solving skill. As discussed, strategic flexibility (i.e., the ability to solve equation in multiple ways, preferably with the most efficient method) is an important aspect of skill and understanding in many domains, including algebra (Alibali, 1999; Hatano & Inagaki, 1986; Star & Rittle-Johnson, 2008; Star & Seifert, 2006) and can probably not be mastered with a tutor that offers only limited freedom. The current study strongly suggests that allowing strategy freedom in an ITS in itself is not enough to improve students' strategic flexibility. Therefore, an ITS geared toward fostering strategic flexibility will need to do more than *allow* multiple solutions. It may need to actively support students in exploring and comparing multiple solutions (cf. Ainsworth et al., 1998; Alibali, 1999; Star & Rittle-Johnson, 2008), or it may have to ask students to solve the same problem multiple times, with different strategies (cf. Ainsworth et al., 1998).

Our investigations of strategy freedom during tutored problem solving have some bearing on the bigger debate in the educational psychology literature on whether discovery-oriented approaches or more structured (or direct) instructional approaches are more educationally effective (Dean & Kuhn, 2006; Koedinger & Aleven, 2007). Several researchers claim that students learn with greater understanding when they discover their own procedures instead of only adopting instructed procedures (Von Glasersfeld, 1995). Others claim that direct instruction is better, partly because discovery learning can overload working memory, or because discovery learning is inefficient, or does not lead to good solutions in the first place (Klahr & Nigam, 2004). Much empirical research is still needed to settle this question (Kirschner, Sweller, & Clark, 2006). The current work does not address this question head on, because it focuses on comparing different conditions that all fall within the realm of tutored problem solving. That is, the current study does not have a discovery learning condition; even the free-est condition is relatively "structured." Nonetheless, a key implication of the current work, namely, that students need some help in taking advantage of freedom offered during tutored problem solving, appears to mirror findings in the literature on discovery learning, namely, that pure discovery learning is challenging for students, and that certain carefully designed forms of guidance during discovery learning are helpful, such as "structured invention activities" (Kapur, 2008; Roll, Aleven, & Koedinger, 2010; Schwartz, Chase, Oppezzo, & Chin, 2011; Schwartz & Martin, 2004), or tutoring support in the context of learning with simulations (Borek, McLaren, Karabinos, & Yaron, 2009; de Jong & van Joolingen, 1998) or experimental design activities (Siler, Klahr, & Price, 2012).

We see several key contributions for the current work. First, to the best of our knowledge, the work is the first to investigate the educational value of offering strategy freedom during (tutored) problem-solving practice with an ITS. Strategy freedom is often assumed to be a necessary element in ITS, but this assumption has not been empirically tested. We found substantial evidence to suggest that supporting minor strategy variations is indeed a good thing to do, since it helps reduce the occurrence of valid strategy variations marked wrong by the system, even if in the current study it was not shown to enhance students' learning results. We found no evidence that supporting major strategy variations is equally useful, but the current work should not be the last word on this issue, as discussed.

A second contribution of the work is the insight it gives in students' strategy use during algebraic problem solving, based on analysis of tutor log data. We are not aware of prior work in the ITS or learning sciences literature that provides this kind of analysis. The surprising insight from this analysis is that the range of students' strategies is quite limited. As mentioned, to a degree, this finding may reflect students' specific instructional background (e.g., the fact that they may have learned the standard strategy), but it may also reflect a general reluctance to explore, possibly due to prior instruction not supportive of exploration.

A third contribution of the work is that it draws attention to an issue that so far has not been discussed much in the literature on ITS and advanced learning technologies, namely, the pros and cons of supporting multiple paths within tutor problems. As mentioned in the introduction, ITS developers and theorists (e.g., Koedinger & Corbett, 2006; VanLehn, 2006) often assume that this capability is desirable, and have long built it into their systems. The current work is to the best of our knowledge the first that provides empirical support for that important assumption. The work also provides some nuance with respect to theories about desirable (system) behaviors in ITS, in particular the notion that ITS are effective because they provide support at the step level (i.e., they break down problems into steps and help students with the steps), rather than at the problem level (e.g., VanLehn, 2006, 2011). The current work suggests that such step-by-step guidance should be given in a flexible manner, while offering freedom to students to pursue different strategies within a given problem. In other words, the work suggests that the effectiveness of ITS derives at least in part from their ability to provide step-by-step guidance with respect to multiple paths through a given tutor problem. The work therefore validates a source architectural complexity in ITS, namely, the ability to support multiple paths. It is important to note that the architectural features needed to support *minor* strategy variations within tutored problem solving (i.e., within an ITS) are no simpler than those needed to support *major* strategy variations. The knowledge represented within the architecture may be simpler (e.g., fewer rules needed, fewer paths in behavior graphs, etc.), but the fundamental tutoring approaches embedded in these systems (model-tracing (Koedinger & Corbett, 2006), example-tracing (Aleven, McLaren, Sewall, & Koedinger, 2009), constraint matching (Mitrovic et al., 2002)) must be capable of interpreting student behavior with respect to alternative paths.

A recommendation that follows from this work is that ITS designers consider gradually introducing complexity into their tutors, rather than assuming from the start that a fully flexible tutor is preferred. Student prior knowledge and the educational goals of the ITS should be taken into account when deciding how much freedom to support. Early pilot studies can start with versions that may be simpler than commonly thought, and introduce complexity gradually. It may even be reasonable to use relatively simple versions in classrooms (though perhaps not the very simplest version) in the early stages of learning a complex skill, in line with Star's (2005) observations. Further, researchers should focus on investigating how ITS can support the development of strategic flexibility. For example, ITS may need to be designed so that students compare and contrast strategies and learn to select the most appropriate ones for any given problem. In pursuing these goals, it will be interesting to adopt and apply the tutor log analysis technique introduced in this work, namely, to analyze how much students use the freedom offered by the ITS. The use of open data repositories with log analysis tools such as the DataShop (Koedinger et al., 2008) will be quite helpful. Knowing more about students' strategy use in tutors may ultimately lead to better tutors, tutors aimed at supporting flexible strategy acquisition.

7. Conclusion

We investigate the value of supporting strategy freedom in an ITS, systems that provide step-by-step guidance during problem-solving activities. Greater strategy freedom requires a more complex systems architecture than supporting only a single solution path through a tutor problem. It is therefore important to ask whether the greater architectural complexity has benefits with respect to student learning and motivation.

In our study, an ITS helped students improve their equation-solving skill. However, allowing minor or major strategy variations did not make a difference in learning gain, motivation or perceived strategy freedom, compared to strictly enforcing a standard strategy with which students were familiar, without allowing any variations. Analysis of tutor log data turned out to be a useful tool for analyzing students' strategy use. It revealed that students tended to use the standard strategy and its minor variations, but that major strategy variations were very rare. Thus in complex problem-solving activities, merely giving student the freedom to use whatever strategy they like does not necessarily pay off; it does not appear to lead to exploration of strategies other than those on which students' prior instruction presumably has focused.

On the basis of these results, we conclude that allowing *minor* strategy variations during tutored problem solving is useful, simply because students use them frequently. If minor strategy variations are not allowed, many student attempts at solving steps would be marked as errors, even though they are actually valid variations. It is possible (and we would be concerned) that such unnecessarily flagged errors would be detrimental to learning in other populations of students. On the other hand, the results do not provide support for a tutor that supports a wide range of major strategy variations during tutored problem solving, at least not without additional support for helping students make use of this freedom. The students in the study did not use major strategy variations, quite possibly as a result of having learned (prior to the experiment) the standard strategy as a safe and sound way to solve linear equations. They did not feel restricted by the stricter versions of the tutor. With respect to the goal of learning to solve linear equations reliably, then, it may be that a tutor that supports a standard strategy with minor variations is sufficient. However, a multi-strategy tutor may still be quite useful, possibly even necessary, for the more ambitious and desirable goal (not attained in the current experiment) of supporting students in acquiring strategic flexibility.

The study further suggests that allowing strategy freedom in an ITS (and perhaps during problem-solving practice more generally) is not sufficient to improve strategic flexibility. In order to do so, the ITS may need to actively support students in exploring and comparing multiple solutions (cf. Alibali, 1999; Star & Rittle-Johnson, 2008). For example, the tutor can give the student instructions and hints for more optimal strategies, or involve them in comparing solutions. To support these kinds of activities, a multi-strategy tutor could well be very useful or even necessary.

In sum, the experiment provided support for the notion that ITS should support multiple solutions strategies during tutored problem solving, but also pointed to the need for further research into how they can help students make optimal use of this freedom.

Acknowledgments

Funding for this research was provided by the Pittsburgh Science of Learning Center which is funded by the NSF award No. SBE-0836012. We would like to thank Ken Koedinger, Agnes Pavlovich, Kenneth Labuskes, and Jesica Kalka for their support and the students, teachers and principals for participating in the study. Finally, Brett Leber, Jonathan Sewall, Octav Popescu, Alida Skogsholm, and Martin van Velsen for their technical support.

References

Ainsworth, S., Wood, D. J., & O'Malley, C. (1998). There is more than one way to solve a problem: evaluating a learning environment that supports the development of children's multiplication skills. *Learning and Instruction, 8*(2), 141–157.
Aleven, V. (2010). Rule-based cognitive modeling for intelligent tutoring systems. In R. Nkambou, J. Bourdeau, & R. Mizoguchi (Eds.), *Advances in intelligent tutoring systems* (pp. 33–62). Berlin: Springer.
Aleven, V., & Koedinger, K. R. (2002). An effective meta-cognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*(2), 147–179.
Aleven, V., McLaren, B. M., & Sewall, J. (2009). Scaling up programming by demonstration for intelligent tutoring systems development: an open-access website for middle-school mathematics learning. *IEEE Transactions on Learning Technologies, 2*(2), 64–78. http://www.computer.org/portal/web/csdl/doi/10.1109/TLT.2009.22.
Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: example-tracing tutors. *International Journal of Artificial Intelligence in Education, 19*(2), 105–154.
Alibali, M. W. (1999). How children change their minds: strategy change can be gradual or abrupt. *Developmental Psychology, 35*(1), 127–145.
Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: lessons learned. *The Journal of the Learning Sciences, 2*(2), 167–207.
Arroyo, I., Woolf, B. P., & Beal, C. R. (2006). Addressing cognitive differences and gender during problem solving. *Technology, Instruction, Cognition, and Learning, 4*, 31–63.
Beal, C. R., Walles, R., Arroyo, I., & Woolf, B. P. (2007). Online tutoring for math achievement: a controlled evaluation. *Journal of Interactive Online Learning, 6*, 43–55.
Benson, J., Dodge, S., Dodge, W., Hamberg, C., Milauskas, G., & Rukin, R. (1991). *Teachers edition algebra 1, an integrated approach*. McDougal, Littell Mathematics.
Borek, A., McLaren, B. M., Karabinos, M., & Yaron, D. (2009). How much assistance is helpful to students in discovery learning? In U. Cress, V. Dimitrova, & M. Specht (Eds.), *Proceedings of the fourth European conference on technology enhanced learning, learning in the synergy of multiple disciplines (EC-TEL 2009), LNCS 5794*. September/October 2009, Nice, France (pp. 391–404) Springer-Verlag Berlin Heidelberg.
Cen, H., Koedinger, K. R., & Junker, B. (2007). Is over practice necessary? – improving learning efficiency with the cognitive tutor through educational data mining. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Proceedings of 13th international conference on artificial intelligence in education (AIED2007)* (pp. 511–518). Amsterdam: IOS Press.
Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715–730.
Dean, D., Jr., & Kuhn, D. (2006). *Direct instruction vs. discovery: The long view*. Wiley Inter Science. www.interscience.wiley.com.
Feng, M., Heffernan, N. T., & Koedinger, K. R. (2009). Addressing the assessment challenge in an intelligent tutoring system that tutors as it assesses. *The Journal of User Modeling and User-Adapted Interaction, 19*, 243–266.
Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions in Education, 48*, 612–618.
Hatano, G., & Inagaki, I. (1986). Two courses of expertise. In H. A. H. Stevenson, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262–272). New York: Freeman.
Holt, Rinehart, & Winston. (2007). *Holt mathematics teachers edition, course 2*.
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.
Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379–424.
J. Kilpatrick, J. Swafford, & B. Findell (Eds.), (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: Academy Press, Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education.
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.
Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: effects of direct instruction and discovery learning. *Psychological Science, 15*, 661–667.
Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239–264.
Koedinger, K. R., Aleven, V., Heffernan, N., McLaren, B., & Hockenberry, M. (2004). Opening the door to non-programmers: authoring intelligent tutor behavior by demonstration. In J. C. Lester, R. M. Vicario, & F. Paraguaçu (Eds.), *Proceedings of seventh international conference on intelligent tutoring systems, ITS 2004* (pp. 162–174). Berlin: Springer Verlag.
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*(1), 30–43.
Koedinger, K. R., Baker, R., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2011). A data repository for the EDM community: the PSLC DataShop. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J.d. Baker (Eds.), *Handbook of educational data mining* (pp. 43–55). Boca Raton, FL: CRC Press.
Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: technology bringing learning sciences to the classroom. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–78). New York: Cambridge University Press.
Koedinger, K., Cunningham, K., Skogsholm, A., & Leber, B. (2008). An open repository and analysis tools for fine-grained, longitudinal learner data. In R. S. J.d. Baker, T. Barnes, & J. E. Beck (Eds.), *Educational data mining 2008: 1st International conference on educational data mining, proceedings*. Montreal, Quebec, Canada. June 20–21, 2008 (pp. 157–166).
Koedinger, K. R., & Mathan, S. (2004). Distinguishing qualitatively different kinds of learning using log files and learning curves. In the working notes of the ITS2004 workshop on analyzing student–tutor interaction logs to improve educational outcomes.
Le, N. T., & Pinkwart, N. (2011). Adding weights to constraints in intelligent tutoring systems: does it improve the error diagnosis? In C. Delgado Kloos, D. Gillet, R. M. Crespo Garcia, F. Wild, & M. Wolpers (Eds.), *Towards ubiquitous learning: 6th European conference on technology-enhanced learning, EC-TEL 2011* (pp. 233–247) Berlin: Springer.
Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum Associates.
Mitrovic, T., Martin, B., & Mayo, M. (2002). Using evaluation to shape ITS design: results and experiences with SQL-tutor. *User Modeling and User-Adapted Interaction, 12*(2–3), 243–279.
Mitrovic, A., Mayo, M., Suraweera, P., & Martin, B. (2001). Constraint-based tutors: a success story. In L. Monostori, J. Vancza, & M. Ali (Eds.), *Proc. IEA/AIE-2001* (pp. 931–940). Budapest: Springer-Verlag Berlin.
Murray, T. (2003). An overview of intelligent tutoring system authoring tools: updated analysis of the state of the art. In T. Murray, S. Blessing, & S. Ainsworth (Eds.), *Authoring tools for advanced learning environments* (pp. 491–544). Dordrecht, the Netherlands: Kluwer Academic Publishers.
NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology, 102*(4), 896–915.
Ritter, S., Kulikowich, J., Lei, P., McGuire, C., & Morgan, P. (2007). What evidence matters? A randomized field trial of Cognitive Tutor® algebra I. In T. Hirashima, H. U. Hoppe, & S. Shwu-Ching Young (Eds.), *Supporting learning flow through integrative technologies* (pp. 13–20). Netherlands: IOS Press.
Roll, I., Aleven, V., & Koedinger, K. R. (2010). The invention lab: using a hybrid of model tracing and constraint-based modeling to offer intelligent support in inquiry environments. In V. Aleven, J. Kay, & J. Mostow (Eds.). *Proceedings of the 10th international conference on intelligent tutoring systems, ITS 2010, Vol. 1* (pp. 115–124). Berlin: Springer.
Ryan, R. M. (1982). Control and information in the intrapersonal sphere: an extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*, 450–461.
Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: a review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology, 45*, 736–750.
Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: the effects of telling first on learning and transfer. *Journal of Education Psychology, 103*(4), 759–775.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: the hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129–184.
Siler, S. A., Klahr, D., & Price, N. (2012). Investigating the mechanisms of learning from a constrained preparation for future learning activity. *Instructional Science*, . http://dx.doi.org/10.1007/s11251-012-9224-7.
Stamper, J., Koedinger, K. R., Baker, R., Skogsholm, A., Leber, B., Demi, S., et al. (2011). Managing the educational dataset lifecycle with DataShop. In J. Kay, S. Bull, G. Biswas, & T. Mitrovic (Eds.), *Proceeding of the 15th international conference on artificial intelligence in education (AIED2011)* (pp. 557–559). Berlin: Springer.
Star, J. R. (2005). Students' use of standard algorithms for solving linear equations. In G. Lloyd, M. Wilson, J. Wilkins, & S. Behm (Eds.), *Proceedings of the twenty-seventh annual meeting of the North American chapter of the International Group for the Psychology of Mathematics Education*. Blacksburg, VA: Virginia Tech University.
Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: the case of equation solving. *Learning and Instruction, 18*, 565–579.
Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology, 31*, 280–300.
Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press.
VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16*, 227–265.
VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221.
VanLehn, K., Lynch, C., Schultz, K., Shapiro, J. A., Shelby, R. H., Taylor, L., et al. (2005). The Andes physics tutoring system: lessons learned. *International Journal of Artificial Intelligence in Education, 15*(3), 147–204.
Von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning*. Bristol, PA: Falmer Press, Taylor &Francis.
Woolf, B. P. (2009). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. Burlington, MA: Morgan Kaufmann.