Enhancing Robust Learning Through Problem Solving in the Genetics Cognitive Tutor

Albert Corbett (corbett@cmu.edu)
Ben MacLaren (maclaren@andrew.cmu.edu)
Angela Wagner (awagner@cmu.edu)
Human-Computer Interaction Institute, Carnegie Mellon University
Pittsburgh, PA 15213 USA

Linda Kauffman (lk01@andrew.cmu.edu)
Aaron Mitchell (apm1@andrew.cmu.edu)
Department of Biological Sciences, Carnegie Mellon University
Pittsburgh, PA 15213 USA

Ryan S. J. d. Baker (baker2@exchange.tc.columbia.edu)
Department of Human Development,
Columbia University Teachers College, New York, NY 10027

Abstract

In this paper, we examine the impact of three learning activities designed to foster more robust learning in a Genetics Cognitive Tutor module on pedigree analysis problem solving, in an experimental study. The three activities are (1) interleaved worked examples with student explanations; (2) enhanced feedback with tutor-provided explanations of problem solving steps; and (3) explicit scaffolding of the reasoning steps in this abductive process-of-elimination reasoning task. The study included four between-subject conditions, a baseline condition in which students exclusively solved standard problems, and three conditions in which students engaged in one of the new learning activities along with standard problem solving. The scaffolded-reasoning condition was most successful in fostering robust learning, as measured by transfer, retention, and preparation for future learning tests. The enhanced feedback condition, in contrast, yielded the poorest performance on the robust learning measures.

Keywords: Education; Problem solving; Robust Learning; Intelligent Tutors.

Introduction

Problem solving is an essential learning activity across STEM domains. Successful problem solving results in "robust" knowledge: knowledge that is well-grounded in domain knowledge, and as a consequence, is well-retained by students, transfers more readily to related problem situations and prepares students for more successful future learning (Koedinger, Corbett & Perfetti, 2012). One of the well-documented risks in problem solving, across STEM domains, is that students can develop superficial knowledge that fails these tests of robust learning. In particular, when students are not well-prepared for problem solving, they can develop problem solving knowledge which focuses on surface elements in problem situations, formal representations, and features of the learning environment itself (Chang, Koedinger & Lovett, 2003; Chi, Feltovich & Glaser, 1981; Rittle-Johnson & Siegler, 1998).

In this paper we examine how to structure problem solving in an intelligent tutoring system to support robust learning in the domain of genetics. Because of its foundational place in the biological sciences, genetics is a large and growing component of high school biology courses, but it is also viewed as one of the hardest topics in biology by both students and instructors, at the secondary and at the post-secondary level (Tsui & Treagust, 2006). Genetics problem solving is characterized by abductive reasoning. In contrast with deductive hypothesis testing, abductive reasoning starts with a set of observations and reasons backwards to infer properties of the genetic processes that produced the data (e.g., whether a trait is dominant or recessive).

In this paper, we study these issues within a tutor lesson for pedigree analysis in the Genetics Cognitive Tutor (Corbett, Kauffman, MacLaren, Wagner & Jones, 2010), which has been successfully piloted in both high school and college classrooms. Pedigree analysis relies on a complex reasoning process, which nonetheless lends itself to straightforward natural language description. This study examines whether robust learning is supported by a scaffolded reasoning activity prior to conventional problem solving, or by incorporating explicit explanations during problem solving.

The Domain: Pedigree Analysis

Basic pedigree analysis problems pose an interesting challenge both for students and for an intelligent tutoring system. Figure 1 displays a typical pedigree analysis problem, in the Genetics Cognitive Tutor (GCT). This pedigree chart displays four generations in a small family. Females are represented as circles and males as squares. In this family, the founding parents have a daughter affected by a rare genetic trait, as represented by the dark circle. No other family members are affected. The student's task is to determine whether this genetic trait is dominant or recessive, and whether it is X-linked, or transmitted on one of the twenty-two autosomal chromosomes in humans.

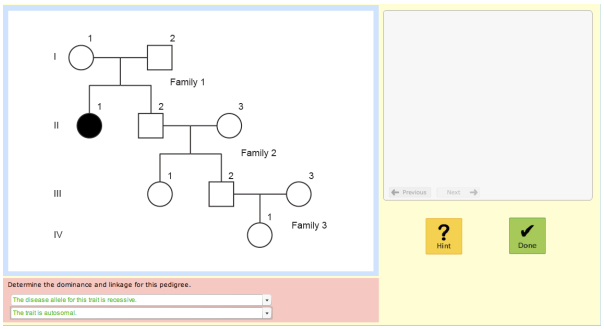

Figure 1. The GCT Interface for Pedigree Analysis.

This appears to be a reasonably simple task; the features of the problem representation are readily interpretable and there are only two problem-solving actions, but the task involves complex abductive reasoning. It requires students to employ their knowledge of genetic transmission to reason by process of elimination. For example, the student can eliminate the possibility that this is a dominant trait, because the daughter must inherit the trait from one or both parents, and if the trait were dominant the parent(s) who have the trait allele would be affected.

This is also an atypical and challenging task for an intelligent tutoring system to effectively support. The task involves a complex multi-step reasoning process, but since there are just two solution steps, there are no natural opportunities (that is, no behavioral correlates of intermediate reasoning steps) for the tutor to provide assistance in the form of feedback and advice along the way. Instead, the task is subject to shallow learning, since students can readily memorize conclusions, (e.g., when an affected daughter has unaffected parents, the trait must be autosomal recessive), without any understanding of the underlying genetics that supports the conclusion. Finally, the task is subject to gaming the system (Baker, Corbett, Koedinger & Wagner, 2004), since each solution step consists of a menu with only three alternatives – "dominant," "recessive," or "cannot be determined," in one case, and "autosomal," "X-linked," or "cannot be determined in the other case." Unmotivated students can readily click through the menu options to find correct answers.

Summative evaluations of Genetics Cognitive Tutor modules are consistent with these risks; pretest-posttest learning gains for basic pedigree analysis are only about half as large as the average gain across all topics (Corbett, et al, 2010).

Pedigree Analysis Learning Activities

In this study we developed and evaluated three Cognitive Tutor activities intended to support robust learning in pedigree analysis problem solving. Two activities integrate explicit reasoning explanations into the conventional problem-solving task - *worked examples*, in which the student explains tutor-generated problem solutions, and *enhanced feedback*, in which the tutor provides explanations for student problem-solving steps. The third activity, in contrast, explicitly scaffolds the intermediate steps in this abductive process-of-elimination task and is designed to precede conventional problem solving. As in Cognitive Tutors more generally (Anderson, Corbett, Koedinger & Pelletier, 1995), in these activities, students receive immediate accuracy feedback on each problem-solving step and can request hints on any problem-solving step.

Interleaved Worked Examples It is well-documented that integrating worked examples with problem solving serves to decrease total learning time and yields improved learning outcomes (Pashler, Bain, Bottge, Graesser, Koedinger, McDaniel & Metcalfe, 2007; Renkl & Atkinson, 2003; Sweller & Cooper, 1985). Recently, several studies have examined the benefits of incorporating worked examples into intelligent tutoring systems (ITSs) for problem solving across a variety of math and science domains, including topics in algebra, geometry, statistics, biology, chemistry and physics (Anthony, 2008; Conati & VanLehn, 2000; Corbett, MacLaren, Wagner, Kauffman, Mitchell, Baker & Gowda, 2011; Mclaren, Lim & Koedinger, 2008; Reed, Corbett, Hoffman, Wagner & MacLaren, 2013; Salden, Aleven, Schwonke & Renkl, 2010; Schwonke, Renkl, Krieg, Wittwer, Aleven & Salden, 2009; Weitz, Salden, Kim & Heffernan, 2010). In these ITS studies, the chief benefit of incorporating worked examples has been to reduce learning time for a fixed set of activities compared to problem solving alone, but unlike the classic worked-example literature, these ITS studies generally do not find that incorporating worked examples leads to more accurate posttest performance than problem solving alone. The exception is Salden, et. al (2010), which found that adaptively fading examples led to some relative improvement on posttest problem solving. Similarly, the evidence that students learn more deeply when worked examples are integrated into ITSs is mixed at best, although Anthony (2008) and Salden, et al (2010) report better retention of problem solving knowledge and Schwonke, et al (2009) found some evidence of greater conceptual transfer in one of two studies.

Pedigree analysis is a promising domain in which to further explore worked examples, since each step in problem solving depends on a complex, but readily describable reasoning process. Figure 2 displays the worked example interface. Each worked example displays a standard pedigree analysis problem and displays the correct dominance and linkage of the trait directly below the pedigree. These examples also identify a key nuclear family in the pedigree and describe the pattern of affected and unaffected individuals in the family that allows the student to identify the dominance and linkage of the trait. Students select entries in the three menus at the bottom of the screen to explain how to determine the dominance and linkage from the pattern, based on their knowledge of genetics transmission.

Feature Focusing We developed a contrasting activity in which the student generates problem solutions and the tutor provides explanations of the student's correct actions, to directly address two characteristics of basic

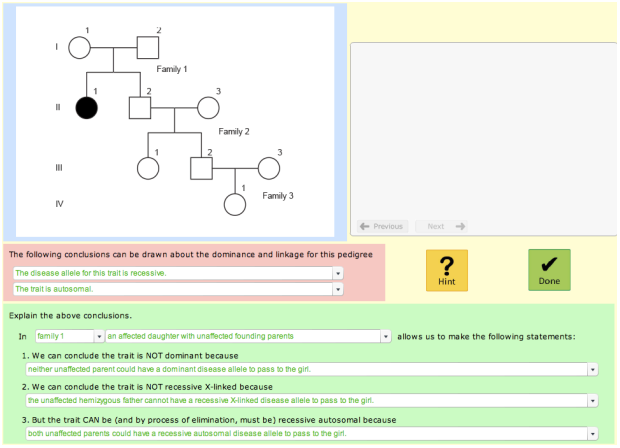

Figure. 2. The GCT Interface at the conclusion of a pedigree analysis worked example.

pedigree analysis. The first is that the tell-tale patterns can be hard to identify. For instance, Figure 3 displays two pedigrees, which look similar, each with four affected males, but the trait on the left is autosomal dominant, while the trait on the right is X-linked recessive.

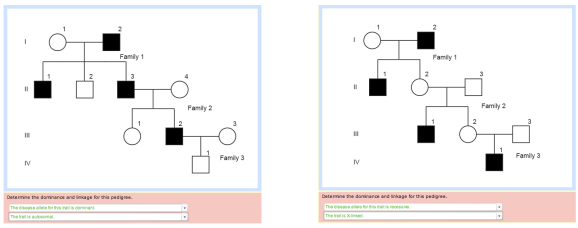

Figure 3. An autosomal dominant pedigree (left) and an X-linked recessive pedigree (right).

The second challenge is that the immediate accuracy feedback generally delivered by Cognitive Tutors (cf. Anderson et al., 1995), is not all that informative in this lesson, since there is a reasonably high probability that the student performed the right action for the wrong reason. Debriefing sessions revealed that students in high school classrooms are aware of the latter risk, and sometimes would like to receive an explanation *after* selecting a correct menu entry, rather than a hint before.

To address these problems, we developed an enhanced feedback interface displayed in Figure 4. The pedigree is initially displayed entirely in black and without any explanatory text. Following each of the two problem-solving steps, the tutor highlights the relevant pattern in the figure, and provides an explanation. In this example, after the student concluded that the trait is recessive, (1) the relevant pattern was highlighted in green, (2) the conclusion was summarized at the top of the screen in green, and (3) an explanation of the conclusion was displayed in green in the window to the right. After the student concluded that the trait is X-linked, the relevant pattern remained highlighted in green, the prior dominance conclusion and explanation were grayed out, the linkage conclusion was summarized in green near the figure, and the linkage explanation was presented in green in the adjoining window.

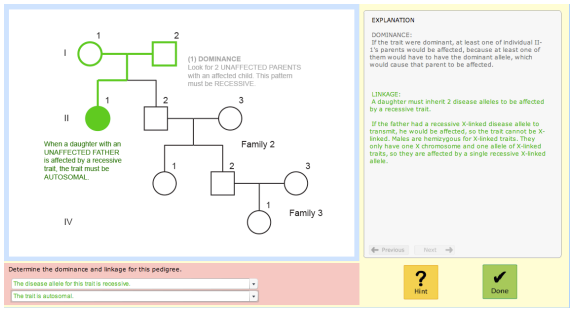

Figure 4. The GCT interface at the end of a pedigree analysis problem with enhanced feedback.

Abductive Reasoning Scaffolds Finally, we developed a problem-solving activity that directly engages students in the reasoning-by-process-of-elimination task. While the other two interventions were integrated with conventional problem solving, this is a separate task that was designed to precede conventional problem solving. Each problem in this task presents the phenotypes of three family members, two parents and a child, as displayed in Figure 5. Immediately to the right, the four possible modes of transmission are listed (autosomal dominant, X-linked dominant, autosomal recessive and X-linked recessive). For each of the four modes, the student enters what the underlying genotype of each of the three family members would have to be, given their respective phenotypes, and under the mode of transmission. (For example, if the trait were autosomal dominant, the two unaffected parents would have to be homozygous recessive, while the affected daughter would have to have a dominant allele.) Then to the far right, the student indicates whether the observed pattern of phenotypes is possible under each mode of transmission, that is, whether the child could inherit its genotype from its parents. (The observed pattern in Figure 5 is impossible for an autosomal dominant trait, since neither parent has a dominant allele to transmit to the daughter.) Finally, at the bottom of the screen the student summarizes which modes of transmission are possible for the observed phenotype pattern.

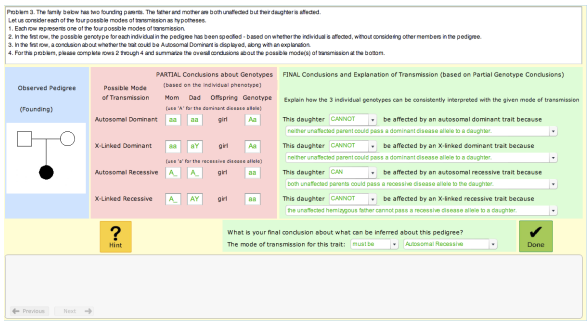

Figure 5. The GCT Pedigree Analysis Scaffolded Reasoning task at the end of a problem.

This study includes four between-subject conditions and evaluates the success of each of these three interventions in supporting the acquisition of problem-solving skills, and robust learning, compared to standard problem solving.

Method

Participants

Sixty-four high school students enrolled in high school biology courses were recruited through newspaper ads and classroom handouts to participate in this study for pay. Students were randomly assigned to one of four between-subject treatment groups.

Procedure

Students participated in two 2.5-hour sessions on consecutive days in a CMU computer lab. In Session 1, students:

- viewed an instructional video and read instructional text on basic pedigree analysis;
- completed a conceptual knowledge pretest and a basic problem-solving pretest;
- completed basic pedigree analysis Cognitive Tutor activities, which differed by condition;
- completed a basic problem-solving test and a transfer problem-solving test.

The second session was devoted to an extended preparation for future learning (PFL) activity, as well as a delayed basic problem-solving test. The PFL task was an advanced carrier-probability pedigree analysis task. Each problem in the task displays a large pedigree chart with five or six generations and students calculate the probabilities that various unaffected individuals in the chart carry a single recessive trait allele. Students:

- read instructional text on carrier probabilities pedigree analysis;
- completed an initial PFL paper-and-pencil test
- completed PFL Cognitive Tutor problems;
- completed a second PFL paper-and-pencil posttest;
- completed a delayed basic problem-solving test

Design

There were four between-subject conditions in the study, defined by students' Cognitive Tutor learning activities in the first session.

- **Basic Problem Solving (PS):** Students completed a set of 78 basic pedigree analysis problems.
- **Enhanced Feedback (EF):** Students completed the same 78 problems as in the PS group, but completed the first 20 with enhanced feedback.
- **Interleaved Worked Examples (WE):** Students completed a problem set with 14 interleaved worked examples and problems to solve, followed by 18 standard problems.
- **Scaffolded Abductive Reasoning (SR):** Students completed six problems in which the abductive reasoning process was explicitly scaffolded as described above, followed by a set of 18 standard problems.

In Session 2, all students completed the same set of activities focused primarily on the PFL task.

Tests

We developed four types of paper-and-pencil tests for the study:

- **Problem Solving Tests**: Three forms were developed. Each form served as the pretest for 1/3 of the students in each condition, the session-1 posttest for 1/3 of the students, and the session-2 delayed test for 1/3 of the students.
- **Conceptual Knowledge Tests**: A conceptual knowledge pretest was developed to evaluate students' knowledge of genetic transmission.
- **Transfer Tests:** A transfer test was developed with two types of problems: one type asked students to solve basic pedigree analysis problems with novel patterns requiring novel reasoning; a second asked students to identify whether family pedigrees were possible or impossible under the four modes of transmission.
- **Preparation for Future Learning (PFL):** Two forms of a PFL problem-solving test were developed. Each form served as the initial test for 1/2 of the students in each condition, and as the second test for 1/2 of the students.

Results

Table 1 displays mean accuracy (percent correct) for the tests administered in the study. The conceptual knowledge (CK) and problem solving (PS1) pretests are displayed to the left, followed by the problem solving posttest (PS2) and the problem-solving learning gain from pretest to posttest (PS2-PS1).

The four robust learning tests follow, including the transfer test (TR); the initial PFL test (PFL1), which preceded the session-2 PFL tutor problems; the second PFL test (PFL2), which followed the GCT PFL problems; and finally the delayed basic problem-solving test (PS3). The final column displays students' change in basic problem-solving accuracy over the retention interval (PS3 – PS2).

Table 1: Student test accuracy (percent correct).

Cond	CK % C	PS1 % C	PS2 % C	PS Gain PS2-PS1	TR % C	PFL1 % C	PFL2 % C	PS3 % C	retention change PS3-PS2
SR	92	47	47	0	54	36	60	53	6
EF	86	41	53	12	44	19	40	49	-4
PS	91	43	48	5	47	34	52	48	0
WE	92	49	56	7	46	31	61	50	-6

Average scores on the Conceptual Knowledge pretest (CK) were quite high, averaging about 90% correct, indicating that students were very familiar with the transmission genetics underlying pedigree analysis. An ANOVA revealed no significant difference among the four conditions on this pretest, $F(3,60) = 1.33$, ns.

Average scores on the Problem Solving pretest (PS1) were much lower, averaging 45% correct. Again, an ANOVA revealed no significant difference among the four conditions on this pretest, $F(3,60) = 1.31$, ns.

Pedigree Analysis Posttest Performance

We performed an ANOVA on the five paper-and-pencil posttest measures of student learning, including the Problem Solving posttest (PS2), and the four robust learning measures: the Transfer test (TR), the Preparation for Future Learning tests (PFL1 & PFL2) and delayed Problem Solving test (PS3). The main effect of condition is not significant, $F(3,60) = 1.26$, ns, but the interaction of condition and test type is significant $F(12,240) = 2.25$, $p < .01$. (The main effect of test type is also significant, $F(4,240) = 28.59$, $p < .001$, but not of particular interest.)

As can be seen in the table, the new enhanced feedback (EF) and scaffolded reasoning (SR) had contrasting impacts. The EF activities yielded the largest problem-solving learning gain, but generally led to the lowest scores on the robust learning tests. In contrast, the new SR activities led to no discernible learning gains from PS1 to PS2, but generally led to the best performance on the robust learning tests.

Basic Problem Solving The new EF condition led to the largest problem-solving learning gains, while the new SR condition led to no discernible learning gains. However, in an ANOVA on the PS gain displayed in Table 1, the effect of condition was not significant, $F(3,60)=1.74$, $p<.17$.

Robust Learning The SR condition generally outperformed the familiar PS and WE conditions, which in turn outperformed the EF condition on the robust learning measures. The difference is fairly pronounced on the transfer task, and in the retention change scores, where the SR condition is the only condition that displays a small increase in scores over the retention interval. We performed an ANOVA on the transfer test, two PFL tests and the retention change scores, and the effect of condition is significant, $F(3,60) = 2.80$, $p < .05$. The interaction of condition and test measure is not significant.

We performed an ANOVA on each of these four robust learning measures separately and condition was significant only for the retention change measure, $F(3,60)=3.41$. $p < .05$, where the SR group is the only one which shows any sign of improving on basic problem solving by virtue of completing the intervening Cognitive Tutor PFL task.

Tutor Performance

Session 1 Total Time Table 2 displays the total time that the students in the four conditions spent on session-1 GCT pedigree analysis learning activities. The session-1 tutor activities were designed to hold learning time constant. As can be seen, average time was reasonably constant across conditions, ranging from about 24 to about 27 minutes. We performed an ANOVA on session 1 time on task, and condition was not reliable, $F(3,60)=0.74$, ns.

Table 2: Student performance on GCT activities in Session 1 and Session 2

	Session 1	Session 2 (PFL GCT)		
Cond	Min.	Min.	%C	%hints
SR	27.4	25.4	58	19
EF	24.9	31.2	41	41
PS	26.0	26.6	49	32
WE	23.7	26.1	55	26

Session 2 PFL Tutor Problems Student performance in the session-2 carrier probabilities GCT task provides an additional PFL measure with respect to the four session-1 learning activities. All students completed the same set of 14 carrier probability problems in the second session. Table 2 displays the average time to complete the problems, student accuracy (the percentage of problem-solving steps on which students' first action was correct), and help requests (the percentage of steps on which a student requested a tutor hint). The students in the SR condition were the most successful in session 2, responding most accurately, while requiring the least time, and least assistance. In contrast, students in the EF condition performed least successfully on all three measures. In three ANOVAs, the main effect of condition is significant for accuracy, $F(3,60)=2.77$, $p<.05$, and for hint requests, $F(3,60)=3.55$, $p<.05$, but not significant for total time, $F(3,60)=1.86$, ns.

Summary and Discussion

Among the three new GCT tasks, the scaffolded reasoning task was the most successful in preparing students for more robust learning in problem solving. The SR combination of a scaffolded reasoning task, in conjunction with a single set of conventional problems, yielded the most robust understanding of pedigree analysis, as measured by transfer, preparation for future learning, and retention of problem-solving skill.

However, design work remains to be done, since the scaffolded reasoning task did not prepare students well for conventional problem solving. Despite their robust learning, students in this condition performed surprisingly poorly on the problem solving posttest, displaying no learning gains.

A more promising design may be to insert the worked example task between the scaffolded reasoning task and conventional unassisted problem solving, to provide students the opportunity to reflect on, and describe how to apply their abductive reasoning skills in

the full problem-solving task. While students in the WE condition did not perform discriminably better than students in the baseline PS condition across the board, there was at least a trend for the WE students to outperform the PS students on the PFL measures.

Finally, the newly designed enhanced-feedback problem solving condition was disappointing. There was a modest and non-significant trend for the EF condition to yield larger learning gains on the problem solving test, but the EF condition led to generally poorer performance on measures of robust learning, especially the PFL test and tutor activities. This may indicate that, to the extent there is a benefit of the enhanced feedback, students are learning to identify the key patterns in pedigrees and to associate them with the corresponding conclusions, but are not developing an understanding of the underlying reasoning. Again, inserting the interleaved worked example activity between the EF task and conventional problem solving might help students build more effectively on any benefits of the EF condition.

Acknowledgments

This research was supported by the Institute of Education Sciences award R305A090549 "Promoting Robust Understanding of Genetics with a Cognitive Tutor that Integrates Conceptual Learning with Problem Solving."

References

Anderson, J.R., Corbett, A.T., Koedinger, K.R. & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, *4*, 167-207.

Anthony, L. (2008). *Developing handwriting-based intelligent tutors to enhance mathematics learning.* Unpub. doctoral dissertation, Carnegie Mellon University.

Baker, R.S., Corbett, A.T., Koedinger, K.R. & Wagner, A.Z. (2004). Off-task behavior in the cognitive tutor classroom: When Students "Game the System". *Proceedings of ACM CHI'2004 Conference on Human Factors in Computing Systems*, 383-390.

Chang, N. M., Koedinger, K. R., & Lovett, M. C. (2003). Learning spurious correlations instead of deeper relations. *Proceedings of the 25th Cognitive Science Society*, 228-233.

Chi, M.T.H, Feltovich, P.J. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *18*, 439-477.

Conati, C. & VanLehn, K. (2000). Further results from the evaluation of an intelligent tutor to coach self-explanation. In the *Proceedings of the 5th Internationational Conference on Intelligen Tutoring Systems*, 304-313.

Corbett, A., Kauffman, L., MacLaren, B., Wagner, A., & Jones, E. (2010). A Cognitive Tutor for genetics problem solving: Learning gains and student modeling. *Journal of Educational Computing Research*, *42*, 219-239.

Corbett, A., MacLaren, B., Wagner, A., Kauffman, L., Mitchell, A., Baker, R. & Gowda, S. (2011). Preparing students for effective explaining of worked examples in the Genetics Cognitive Tutor. In *Proceedings of the Thirty-third Annual Meeting of the Cognitive Science Society*, 1476-1481.

Koedinger, K.R., Corbett, A.T. & Perfetti, C. (2012). The Knowledge-Learning-Instruction (KLI) Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science*, *36*, 757-798.

McLaren, B. M., Lim, S.-J., & Koedinger, K. R. (2008). When is assistance helpful to learning? Results in combining worked examples and intelligent tutoring. In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems* (pp. 677-680).

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing Instruction and Study to Improve Student Learning*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Reed, S.K., Corbett, A., Hoffman, B., Wagner, A., & MacLaren, B. (2013). Effect of worked examples and Cognitive Tutor training on constructing equations. *Instructional Science*, *41*, 1-24.

Renkl, A. & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist, 38*, 15-22.

Rittle-Johnson, B. & Siegler, R.S. (1998). The relation between conceptual and procedural knowledge in learning mathematics: A review. In C. Donlan (Ed.), *The development of mathematical skill* (pp. 75-110). Hove, UK: Psychology Press.

Salden, R. J. C. M., Aleven, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science*, *38*, 289-307.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artifact of lousy control conditions. *Computers in Human Behavior, 25*, 258-266.

Sweller, J. & Cooper, G.A. (1985). The use of worked examples as a substitute for problem solving in learning algebra., *Cognition and Instruction, 2*, 59–89

Tsui, C-Y. & Treagust, D.F. (2006). Understanding genetics: Analysis of secondary students' conceptual status. *Journal of Research in Science Teaching, 44*, 205-235.

Weitz, R., Salden, R. J. C. M., Kim, R. S., & Heffernan, N. T. (2010). Comparing worked examples and tutored problem solving: Pure vs. mixed approaches. In *Proceedings of the Thirty-Second Annual Meeting of the Cognitive Science Society* (pp. 2876-2881).