Differential Impact of Learning Activities Designed to Support Robust Learning in the Genetics Cognitive Tutor

Albert Corbett[1], Ben MacLaren[1], Angela Wagner[1], Linda Kauffman[2], Aaron Mitchell[2], and Ryan S. J. d. Baker[3]

[1]Human-Computer Interaction Inst., Carnegie Mellon Univ., Pittsburgh, PA 15213
[2]Dept. of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, 15213
[3]Dept. of Human Development, Columbia Univ. Teachers College, NY, NY 10027

Abstract. This paper describes two types of Conceptually Grounded Learning Activities designed to foster more robust learning in the Genetics Cognitive Tutor: interleaved worked examples and genetic-process reasoning scaffolds. We report three empirical studies that evaluate the impact of these learning activities on three diverse genetics problem-solving topics in the tutor. We found that interleaved worked examples yielded less basic-skill learning than conventional problem solving, unlike many prior ITS studies of worked examples. We also found preliminary evidence that scaffolded reasoning tasks in conjunction with conventional problem solving leads to more robust understanding than conventional problem solving alone. Implications for the use of contextually grounded learning activities are discussed.

1 Introduction

Problem solving is an essential learning activity across STEM courses. Successful problem solving results in "robust" understanding, grounded in conceptual domain knowledge, that transfers more readily to related problem situations, that is well-retained by students, and that affords more efficient or effective future learning [1]. 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 [2].

This paper describes two types of *Conceptually Grounded Learning Activities* (CGLAs) we have developed to support more robust learning in an intelligent tutoring system for genetics problem solving, and we report the results of three studies that evaluate the impact of these new CGLAs across three problem-solving topics. These two activities are interleaved worked examples, and reasoning scaffolds that link underlying genetics processes with problem solving logic.

Worked Examples. It is well-documented that integrating worked examples with problem solving serves to decrease total learning time and yields improved learning

adfa, p. 1, 2011.

outcomes.[3], [4]. Recently, several studies have examined the benefits of incorporating worked examples into intelligent tutoring systems (ITSs) for problem solving across a variety of STEM domains [5-10]. 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, but unlike the classic worked-example literature, these ITS studies generally do not find that the use of worked examples leads to more accurate posttest performance than problem solving alone. Similarly, the evidence that students learn more deeply when worked examples are integrated into ITSs is mixed at best, although [9] found some evidence of greater conceptual transfer in one of two studies. This paper examines the impact of interleaved worked examples in an ITS for genetics problem solving.

Reasoning Scaffolds. 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 processes that produced the data (e.g., whether a crossover has occurred between two genes during meiosis). This reasoning task is challenging and there is a risk of shallow learning, since students can learn to solve these types of problems algorithmically, based on the formal properties of the problem representations, without reference to the underlying genetics. As a result, we have developed *process modeling* tasks and *solution construction* tasks that are designed to precede standard genetics problem-solving tasks and to ground students' problem-solving knowledge in the underlying genetics prior to problem solving.

These two types of CGLAs have been developed for three topics in an existing Cognitive Tutor for genetics problem solving [11], which has been successfully piloted in both high school and college classrooms. In the following sections we describe these three problem-solving tasks and the new CGLAs, and report results across three studies that examine the impact of these CGLAs on learning.

1.1 The Domain and Learning Activities

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 the post-secondary level. We developed and evaluated CGLAs for three types of genetics problems that represent a diverse range of reasoning tasks: Three-factor crosses, gene interaction, and basic pedigree analysis.

Three-Factor Cross (3FC) Problems. Fig. 1 displays the GCT interface near the end of a *three-factor cross problem*. In these gene-mapping problems, students reason about how crossovers in meiosis reveal the relative positions of genes on a single pair of chromosomes. In each problem, two organisms are crossed, e.g., two fruit flies, and students analyze the relative frequencies of three phenotypic traits among the offspring (displayed in the table on the left of Fig. 1). Each trait is controlled by a single gene and the three genes are located on the same pair of chromosomes. Based on relative frequencies, students infer the order of the three genes on the chromosomes and the relative distance between the three genes.

The basic problem-solving procedure is constant in these problems. The offspring phenotypes fall into four groups and students identify the largest group and smallest group, then identify the middle gene by finding the gene the has switched over relative to the other two, between the two groups. Then students calculate three arithmetic expressions to find the map distances among the pairs of genes.

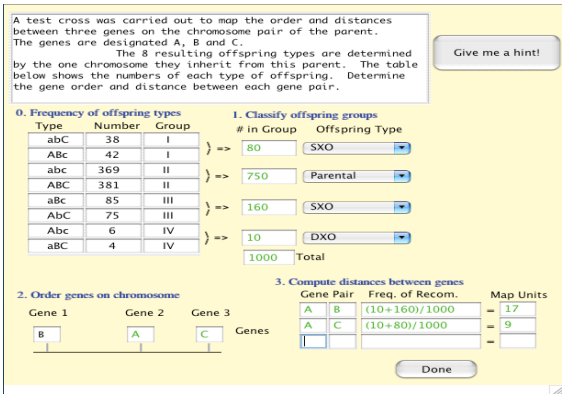

Fig1. The GCT interface near the conclusion of a 3FC problem.

Gene Interaction and Epistasis (GIE) Problems. These problems extend basic principles of Mendelian transmission to traits controlled by two genes. Fig. 2 displays the GCT interface at the end of a problem. Each problem presents three true-breeding strains of an organism and in each of the three columns across the screen, students cross pairs of strains and intercross the resulting offspring. Based on the ratios of the observed offspring phenotypes, students infer the genotypes of the true-breeding strains and infer the genotype of each of the offspring groups in each of the crosses.

Four distinct phenotypes are possible for traits with two genes, but given the different ways two genes can interact, only two or three phenoyptes may be observed. In these problems students reason about both the number of phenotypes, and their relative frequencies, to determine the alleles of each gene that underlie each phenotype.

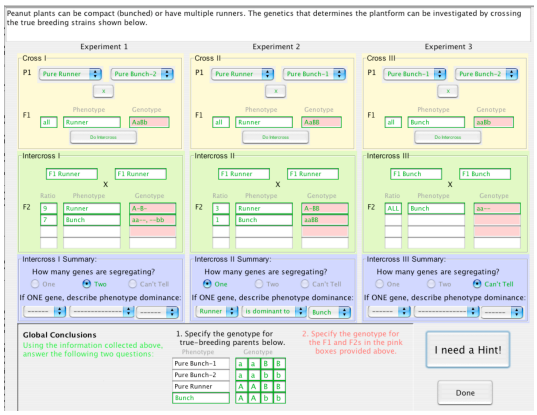

Fig2. The GCT interface at the conclusion of a GIE problem.

Basic Pedigree Analysis (BPA) Problems. Fig. 3 displays a GCT *pedigree analysis* problem. Each problem displays a family tree, including some individuals who are affected by a rare trait. Females are represented as circles and males as squares. In this family, a single male is affected by the rare trait, as represented by the dark square. 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 autosomal chromosomes. The main challenge is to identify the pedigree configurations that identify different transmission modes. Six different conclusions are possible since sometimes the linkage, and/or dominance cannot be determined, and each problem consists of just 2 steps.

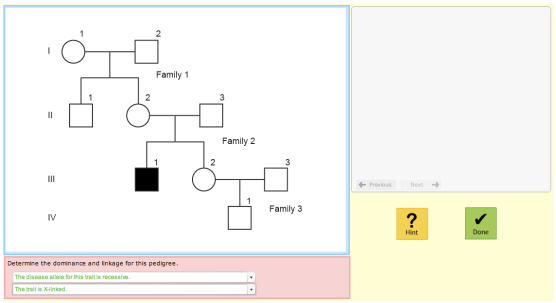

Fig. 3. The GCT Interface for Basic Pedigree Analysis at the end of a problem.

Worked Examples.

There is a substantial risk of shallow learning in genetics problem solving. In pedigree analysis, for example, students can memorize that when two unaffected parents have an affected child, the trait must be recessive, without any understanding of how the properties of the underlying genetics processes support that conclusion. In this project we developed worked example learning activities to explicitly ground students' understanding of problem solutions in the underlying genetic processes. In each case, the worked example interface is constructed around the original problem-solving interface, but includes menus in which students explain the solution steps. As in all Cognitive Tutor activities, students receive accuracy feedback on each menu selection and can ask for help as needed for each menu.

In the 3FC and GIE worked examples, students use two menus to explain each problem-solving step. In the first menu, students describe the features of the empirical evidence that warrant the conclusion, and in the second, they describe why that evidence supports the conclusion based on the underlying genetic processes. The BPA worked examples interface is slightly different, since complex reasoning is packed into just two total steps. In BPA, two menus are used to describe the key pattern in the pedigree that supports both transmission conclusions, and three menus explain how the evidence supports the conclusions based on the underlying genetics. Screenshots of these WE activities, and the SR activities described in the following section, can be viewed at *www.cs.cmu.edu/~genetics/CGLA.html*.

Scaffolded Reasoning. We also developed activities to directly engage students in reasoning about the genetic processes underlying the three types of problem solving tasks. As in all Cognitive Tutor activities, students receive accuracy feedback on each step and can ask for help on each step in these activities.

For the 3FC and GIE topics we developed separate *Forward Modeling* and *Solution Contruction* CGLAs. In abductive reasoning students are given empirical evidence and asked to infer the genetic process that generated the data. In the Forward Modeling tasks, students are given the initial state of a genetic process, and model how the process unfolds to generate empirical data. For example, in 3FC, students are given the ordering and distances among the alleles on the parental chromosomes and model how recombination in meiosis gives rise to offspring phenotypes.

These Forward Modeling activities were coupled with Solution Construction in which students are given both the empirical evidence in a typical problem and the initial state of the underlying process that generated the evidence, and reason through the abductive logic that connects the evidence to the known underlying genetics.

The scaffolded reasoning task was again different for the PA analysis problems. Pilot research showed that students understand the basic transmission genetics that underlie pedigree analysis, so we developed a single Solution Construction activity that scaffolds students' use of that knowledge in solving PA problems. Each problem in this task presents the phenotypes of three family members, two parents and a child. For each of the four possible modes of transmission (autosomal recessive or dominant, X-linked recessive or dominant), the students indicate what the underlying genotype of each family member would have to be, given their phenotypes, and whether the observed pattern of phenotypes is possible under each of the four modes of transmission (i.e., whether the parents have the alleles the child must inherit). Finally, the student summarizes which modes of transmission are possible for the observed phenotype pattern and what final conclusion can be drawn.

2 The Studies

Each study included three conditions defined by the activities described above: a standard problem solving baseline condition (PS), an interleaved worked example condition (WE), and a scaffolded reasoning (SR) condition. Each of the studies included a fourth condition, but these conditions varied across the three studies and are not reported here. A more complete report of all four conditions in the BPA study appears in [12].

The three study procedures varied in specifics, but shared this general structure:

- High school students enrolled in biology courses were recruited through newspaper ads and classroom handouts to participate in the studies.
- The studies were conducted in CMU computer labs and students participated in sessions on two successive days, with each session lasting 2 or 2.5 hours.
- Prior to working with the GCT, students completed a conceptual knowledge pretest and a problem solving pretest.

- After using the GCT, students completed a problem solving posttest and two measures of robust learning: a transfer posttest and a preparation for future (PFL) learning posttest.

Across the three studies, a total of 163 high school students participated in the three treatment conditions reported here. Forty-two students participated in the three-factor cross study; seventy-four students participated in the gene interaction study and forty-seven students participated in the pedigree analysis study. The students participating in each study were randomly assigned to a treatment group.

2.1 Design

The three conditions in each study were defined by students' Cognitive Tutor learning activities in the first study session.

- **Basic Problem Solving (PS):** Students in all three studies only completed standard GCT problems during the first session.
- **Interleaved Worked Examples (WE):** Students completed a problem set in which worked example problems were interleaved with standard problems to solve.
- **Scaffolded Reasoning (SR):** Students completed a block of scaffolded reasoning problems in each study, to prepare them for more robust problem solving. (Students in the 3FC study spent all their time on SR activities in the first session. Students in the GIE and BPA studies spent about 2/3 of their Cognitive Tutor time on SR activities in the first session, followed by standard problem solving)

Students in all conditions within each study concluded their activities with the same block of standard GCT problems to solve.

Students in the 3FC and GIE conditions completed their condition-specific Cognitive Tutor learning activities in the first session. In the second session, students completed a common set of standard Cognitive Tutor problems, followed by the three posttests. The PA problems are intrinsically shorter and students completed all their PA learning activities, including the common block of standard PA problems during the first session. They completed the basic problem solving and transfer posttest the first day and completed their PFL posttest at the beginning of the second session, (followed by additional unrelated Cognitive Tutor activities and tests).

2.2 Tests

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

- **Problem Solving Tests**: Three forms of a basic problem-solving test were developed for each study. Each student received different forms as the pretest and posttest, with each form serving as the pretest for 1/3 of the students and a posttest for a different 1/3 of the students in each condition.
- **Conceptual Knowledge Tests**: A conceptual knowledge pretest was developed for each study to assess students' understanding of the genetic processes that underlie the problem-solving task.

- **Transfer Tests:** A transfer test was developed for each study, challenging students to extend their understanding to novel, related problem situations without further instruction.
- **Preparation for Future Learning (PFL):** A PFL test was developed for each study. Each test presented 2-3 pages of instruction on a new, but related problem-solving task that builds on the genetics knowledge students were acquiring, then asked students to solve problems.

3 Results

Table 1 displays mean accuracy (probability correct) for the tests administered in the three studies. Students' pretest scores are displayed in the two left columns. Average scores on the conceptual knowledge pretest (CK) varied across studies, but varied little across conditions overall. In an ANOVA with study and condition as factors, the main effect of study was significant, $F(2,154) = 168.51$, $p < .01$, but the main effect of condition, and interaction of condition and study were not significant.

Table 1. Student test accuracy (probability correct).

	Pretests p(C)		**Posttests p(C)**			
	CK	PS1	PS2	PS gain	Transfer	PFL
Overall						
SR	0.60	0.32	0.53	0.21	0.51	0.56
WE	0.58	0.26	0.51	0.25	0.45	0.47
PS	0.61	0.27	0.60	0.33	0.46	0.54
3FC						
SR	0.49	0.14	0.43	0.29	0.54	0.67
WE	0.47	0.15	0.53	0.38	0.55	0.58
PS	0.54	0.17	0.65	0.48	0.51	0.75
GIE						
SR	0.40	0.35	0.68	0.33	0.46	0.65
WE	0.35	0.15	0.43	0.28	0.35	0.51
PS	0.37	0.21	0.67	0.46	0.38	0.55
PA						
SR	0.92	0.47	0.47	0.00	0.54	0.36
WE	0.92	0.49	0.56	0.07	0.46	0.31
PS	0.91	0.43	0.48	0.05	0.47	0.34

Average scores on the Problem Solving pretest (PS1) were much lower overall and again varied across studiess. In an ANOVA, the main effect of study was again significant, $F(2,154) = 44.50$, $p < .01$. The main effect of condition was not significant but the interaction of study and condition was significant, $F(4,154) = 3.04$, $p < .05$, so we treat problem solving pretest score as a covariate in all subsequent ANCOVAs.

Posttest Scores. Students' scores on the basic problem solving posttest, pretest-to-posttest learning gains, and two robust learning posttests are displayed in the four data columns at the right of Table1. As can be seen at the top of the table, students in the conventional problem solving condition are performing about 15% better overall on the basic problem-solving test than the other two groups (0.60 vs. 0.52) and the learning gains in the PS group are about 43% larger than in the other two groups (0.33 vs 0.23). However, students in the SR condition score about 13% better on the transfer tests than students in the other two groups (0.51 vs. 0.45), while on the PFL tests the WE group performs about 14% worse than the other two groups (47% vs 55%).

We performed an ANCOVA on the posttest results, with the three tests as a repeated measure, and study and condition as factors. The most important finding is that the interaction of test type (PS2, transfer & PFL) and condition is significant $F(4, 306) = 3.11$, $p < .05$. The main effect of study is also significant, $F(2,153) = 34.94$, $p < .01$; scores in the pedigree analysis study were substantially lower than in the other two studies. (The main effect of study is significant in all subsequent ANCOVAs at the .01 level; and is not reported separately for subsequent ANCOVAS.) The interaction of study and condition is not significant, while the interaction of test type and study is significant, $F(4,306) = 12.16$, $p < .01$ Finally, the three way interaction of test type, study and treatment condition is significant, $F(8,306) = 2.21$, $p < .05$.

Basic Problem Solving Posttests. Further analyses confirm that the PS condition generally led to better acquisition of basic skill than the other conditions. We performed an ANCOVA on the problem-solving posttest alone, and the main effect of condition is significant, $F(2,153) – 4.01$, $p < .05$. The advantage of PS condition is strongest in the 3FC study and weakest in the BPA study, and this interaction of study and condition is marginal, $F(4,153) = 2.14$, $p < .08$.

We also performed an ANCOVA on basic problem solving scores for each pairwise comparison. For the PS and WE conditions, the main effect of condition is significant $F(1,103) = 4.44$, $p < .05$, while the interaction of condition and study is again marginal, $F(2,103)$, 2.38, $p < .10$. For the PS and SR conditions, the main effect of condition is again significant, $F(1,101) = 7.25$, $p < .01$, while the interaction of study and condition is not significant. Finally, comparing the WE and SR conditions, the main effect of condition is not significant, while the interaction of study and condition is marginal, $F(2,101) = 2.41$, $p < .10$.

Robust Learning Posttests. Finally, we performed an ANCOVA with the two robust learning measures as a repeated measure and the main effect of condition was not significant in this analysis. The only significant result in this ANCOVA was the interaction of test type and study, $F(2,152) = 17.52$, $p < .01$,

However, an inspection of the scores in the individual studies in Table 1 show that in five of six robust learning comparisons, performance in the SR group is higher than in the PS group. Performance in the SR condition is also higher than in the WE condition in five of six comparisons. Both of these patterns are marginally significant in a binomial test, $p = .094$.

Session 1 Total Time. Table 2 displays the average time spent on Session 1 GCT learning activities. As can be seen, students completed the GCT tasks more quickly in the PA study than in the other studies. Within each study, however, the session-1 GCT learning activities were designed to hold time on task constant. Across the three studies, students in the PS and WE conditions spent similar amounts of time on the tutor activities, and students in the SR condition spent about 5% more time.

Table 2. Total time for Session 1 GCT learning activities (min.) .

3FC			GIE			BPA		
PS	WE	SR	PS	WE	SR	PS	WE	SR
51	58	53	52	50	55	26	23	27

4 Discussion and Conclusions

In this paper, we present three studies on the use of Cognitively Grounded Learning Activities (CGLAs) in a Cognitive Tutor for Genetics, comparing two CGLAs to a baseline problem-solving condition. While the baseline problem-solving condition led to better acquisition of problem-solving skills than the worked examples or scaffolded reasoning conditions, these studies provide preliminary evidence that reasoning scaffolds that explicitly ground students' problem-solving knowledge in the underlying genetic processes lead to more robust understanding. The benefits in these studies are relatively small; the reasoning scaffolds led to roughly 15% better performance on robust learning measures for GIE and BPA while having little impact for 3FC. Genetic process scaffolding may be less useful in 3FC cross problems because the underlying process in that task is itself relatively simple and the problem-solving procedure is constant across 3FC problems. The underlying genetics processes in the other two domains are more complex and student reasoning varies more across those problems. Therefore grounding student reasoning in the underlying genetic processes may only be helpful when such variation is present

Perhaps the most surprising result across these studies is that interleaved worked examples led to smaller learning gains for basic problem solving than standard problem-solving activities, in contrast with earlier studies of integrated worked examples in intelligent tutoring environments. This may be because the purpose of the explanations in the present studies are somewhat different than in earlier ITS studies. In these genetics studies, the purpose of the explanations is not just to help students refine their understanding of the problem solving steps, but also to ground the problem-solving steps in an underlying causal process model. It may be that this objective is too ambitious to graft onto the problem solving process with explanations; doing so may instead interfere with refining basic problem-solving skills. Hence, the preliminary conclusion from these studies is that students may be more likely to benefit from reasoning about underlying causal models when that experience takes the form of explicit scaffolded-reasoning learning activities that precede problem solving.

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

1. Koedinger, K.R., Corbett, A.T., Perfetti, C.: The Knowledge-Learning-Instruction (KLI) Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science*, *3*6, 757-798 (2012)
2. Bransford, J.D., Brown, A.L., Cocking, R.R. (2000). *How people learn: Brain, mind, experience and school.* Washington, DC: National Academy Press (2000)
3. Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., Metcalfe, J.: *Organizing Instruction and Study to Improve Student Learning*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education (2007)
4. Renkl, A., Atkinson, R. K.: Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist*, *38*, 15-22 (2003)
5. Corbett, A., MacLaren, B., Wagner, A., Kauffman, L., Mitchell, A., Baker, R., Gowda, S.: 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. Austin, TX: Cognitive Science Society (2011)
6. McLaren, B. M., Lim, S.-J., & Koedinger, K. R.: 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*, 677-680 (2008)
7. Reed, S., Corbett, A., Hoffman, B., Wagner, A., MacLaren, B.: Effect of worked examples and Cognitive Tutor training on constructing equations. *Instructional Science*, *41*, 1-24 (2013)
8. Salden, R., Aleven, V., Schwonke, R., & Renkl, A.: The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science*, *38*, 289-307 (2010)
9. Schwonke, R., Renkl, A., Salden, R., & Aleven, V.: Effects of different ratios of worked solution steps and problem solving opportunities on cognitive load and learning outcomes. *Computers in Human Behavior*, *27*, 58-62 (2011)
10. Weitz, R., Salden, R. J. C. M., Kim, R. S., Heffernan, N. T. : Comparing worked examples and tutored problem solving: Pure vs. mixed approaches. *In Proceedings of the Thirty-Second Annual Meeting of the Cognitive Science Society*, 2876-2881 (2010)
11. Corbett, A., Kauffman, L., MacLaren, B., Wagner, A., Jones, E.: A Cognitive Tutor for genetics problem solving: Learning gains and student modeling. *Journal of Educational Computing Research*, *42*, 219-239 (2010)
12. Corbett, A., MacLaren, B., Wagner, A., Kauffman, L., Mitchell, A., Baker, R.: Enhancing robust learning through problem solving in the Genetics Cognitive Tutor. *Proceedings of the Thirty-fifth Annual Meeting of the Cognitive Science Society* (submitted)