Enhancement of Student Learning Through the Use of a Hinting Computer e-Learning System and Comparison With Human Teachers*

Pedro J. Muñoz-Merino, *Member, IEEE*, Carlos Delgado Kloos, *Senior Member, IEEE*, and Mario Muñoz-Organero, *Member, IEEE*

Abstract: **This paper reports the results of an experiment in a Computer Architecture Laboratory course classroom session, in which students were divided into two groups for interaction both with a hinting e-learning system and with human teachers generating hints. The results show that there were high learning gains for both groups, demonstrating the effec-tiveness of the human teachers as well as of the computer-based hinting e-learning system even without the use of adaptive and personalization ca-pabilities. In addition, in the worst case, the difference in favor of human teachers (with a low student-to-teacher ratio of 13.5 students per teacher) would not be significant with respect to the e-learning system, so the com-puter-based system can replace teachers without a significant loss of effec-tiveness.**

Index Terms: **Assessment, computers, e-learning, evaluation, hints.**

I. Introduction and Related Work

The concept of hinting in problem solving can be defined as the pro-vision of some task that helps the user in the resolution of an initial problem. [1] gives a review of the provision of hints by human and com-puter tutors. Some examples of hinting tutors are Andes [2], SIETTE [3], PACT [4], or AgentX [5]. A hinting software module [6] was imple-mented within XTutor (http://icampus.mit.edu/xTutor/). This hinting module tries to combine most of the nonadaptive functionality present in state-of-the-art hinting systems with other functionality based on new ideas, such as a new scoring method with penalties or rewards for hint resolution and for viewing hints, previous meta-information about the hints, a maximum limit of hints to select, or an undefined level of nested hints between problems.

[7] reports positive results for the use of the PACT tutor in com-parison with traditional teaching. In [8], three groups of students were compared: those who learned collaboratively in groups, those who were provided with hints, and those who learned individually without hints. The first two groups were shown to perform better. [9] showed the effectiveness of computer-delivered hints in developing problem-solving abilities. [6] shows that better results were obtained with the XTutor hinting module than with an assessment e-learning system without hints. All of these studies concluded e-learning sys-tems that include hints are more effective than those learning situations in which hints are not available. An innovative aspect of the work reported here, which differentiates from the literature reported here, is that it makes a direct comparison between computer-delivered hints and those delivered by teachers in a typical classroom session, who are encouraged to give hints to students so as to maximize their learning gains.

Human teachers have usually been considered as being very good at providing hints. The study reported in [10] suggested that the reason for the success of a human teacher with respect to a computer is that humans do not provide the solution directly. [11] judges the quality of a computer-based tutor by assessing the similarity between the number of hints generated by the tutor and by a human. However, the conditions of the research reported here are different: The hinting contents of the tutor are initially set by teachers, and there is a comparison of learning gains.

This work does not include any adaptation capabilities because the aim is to know the limitations of hinting capabilities without adapta-tion. The goals of this research are to measure the learning gains pro-duced by the hinting e-learning system during a classroom session and by human teachers providing hints and to compare both.

II. Outline of the Hinting e-Learning System

Fig. 1 shows the initial phase of the interaction within the hinting system used for a sample problem. the student is set an exercise on makefile. The student can request a hint, upon which the information shown in Fig. 2 appears; this is only meta-information about the initial problem. Next, the student can choose to view one or more hints (there is not a maximum limit in this case, but there are scoring penalties) with the "plus" buttons; in that case, other problems would appear as hints. Within the hints themselves, further explanatory hints can also be provided. There are also penalties for incorrect attempts. A complete description of the different features of the hinting module is available in [6]. The exercises with hints, like that shown, were designed with the help of an implemented authoring tool [12].

III. Methodology

This study took place in a Computer Architecture Laboratory class-room session that lasted for 2 h. Students worked individually with an assigned computer. The topics covered were shell script, makefile, semaphores, and the FAT file system.

Students were divided into two experimental groups: group A (55 students) that interacted with the tutor system generating hints and group B (27 students) that interacted with the teachers providing hints. The same 16 root exercises were provided for both groups. There was a ratio of 13.5 students per teacher in group B.

Both experimental groups performed a pre-test, then underwent their interaction with the hinting system (group A) or with human teachers (group B), followed by a post-test. The pre-test and post-test each con-sisted of eight questions and were designed to have the same difficulty level; no questions were repeated. Eight questions can be an adequate predictor of student knowledge in this case because the experiment was focused on evaluating only a reduced set of selected key concepts. The pre-test and post-test each took 10 min, while the interaction with the

The authors are with the Carlos III of Madrid University, 28911 Leganes, Madrid, Spain (e-mail: pedmume@it.uc3m.es; cdk@it.uc3m.es; munozm@it.uc3m.es).

Suppose the current directory contains the files main.c, part1.c, part2.c, makefile, clean where main.c uses the functions defined in part1.c and part2.c and the .c files can be compiled without errors. Select all the true statements related to this makefile, which code is provided next (each incorrect attempt subtracts 3 points)

Línea	Código
1	`myprogram: main.o part1.o part2.o`
2	`        gcc -o myprogram main.o part1.o part2.o`
3	
4	`part1.o: part1.c`
5	`        gcc -c part1.c`
6	
7	`part2.o: part2.c`
8	`        gcc -c part2.c`
9	
10	`main.o: main.c`
11	`        gcc -c main.c`
12	
13	`clean:`
14	`        rm -f myprogram main.o part1.o part2.o`

☐ A single call of 'make myprogram' leads to the creation of an executable file ☐ A single call of make clean does produce errors ☐ Lines 4 to 11 can be substituted by a convenient implicit rule and it is equivalent ☐ Line 2 can be substitued by 'gcc -o $?' and it is equivalent
SCORE: 0.0/10.0

Check

Fig. 1. Exercise with hints executing within XTutor (first phase of the interaction).

hinting system or with human teachers lasted for 50 min. The rest of the time was devoted to filling in a survey and to exchanging the groups' interaction types to ensure fairness.

There were no adaptation capabilities to select the different hinting techniques, but each exercise had a limited number of established hinting techniques. Some of the hinting techniques were set randomly at the beginning, but in such a way that each student would encounter the same number of different opposite hinting techniques while performing the problems and that each problem would be addressed with the same number of different opposite hinting techniques by the various students attempting to solve it.

IV. RESULTS AND DISCUSSION

Table I shows the students' grades (scale from 0 to 10) for both experimental groups in the pre-test and post-test. Applying the t-test to the students who interacted with the computer hinting system, there is a statistically significant difference in learning gain, measuring the difference between the post-test and pre-test ($t = 7.450, n = 102, p = 0.000$). In order to quantify the increase of the students' knowledge, applying the t-test with a 90% confidence interval, the learning gain would be between [2.06, 3.25]. So with a 95% probability, learning gain would be greater than 2.06.

On the other hand, the interaction of students with human teachers providing hints led to a significant difference in learning gain between the post-test and the pre-test according to the t-test ($t = 5.147, n = 52, p = 0.000$). Applying the t-test again with a 90% probability for confidence intervals, the learning gain would be between [1.69, 3.31].

The standard errors for the human teachers are greater than those for the hinting system. This difference can be explained by there being twice the number of students interacting with the hinting tutor. The pre-test standard deviation was high for both groups, which raises the question of the heterogeneity of the students' initial knowledge levels. The standard deviations and standard errors of the post-test are lower than those for the pre-test, which could be explained as a homogeneity

effect on the students' final knowledge levels, produced both by the teachers and by the hinting tutor.

These results corroborate that both methods (the computer hinting tutor and human tutors giving hints) were very effective for the acquisition of knowledge, and the learning gains obtained were impressively high. Indeed, such a high increase in knowledge can seem surprising, after only 50 min of interaction (with the human teachers or with the computer-based system). However, these impressive learning gains should be considered in the context of the experiment: the session was focused on a set of only a few selected key concepts in shell script, makefile, semaphores, and FAT; students were properly motivated at the beginning; and the learning gains were measured immediately, while the experience was still fresh in the students' minds.

Fig. 3 shows the means of the knowledge increase (mean of the different students' differences between post-test and pre-test) for each group: 2.65 for the students who interacted with the hinting tutor, with a standard deviation of 1.35 and a standard error of 0.19; and a mean of 2.50 for the students who interacted with the human teachers, with a standard deviation of 1.30 and a standard error of 0.25.

Applying the t-test for the knowledge increase means, there is not a statistically significant difference between both mean values ($t = 0.457, n = 77, p = 0.649$), and the confidence interval for differences of knowledge increase between the hinting computer-based system and human tutors is within the interval $[-0.38, 0.70]$ with a 90% probability. Therefore, the difference in favor of human teachers would not be greater than 0.38 with a 95% probability. This amount can be considered insignificant, compared to the total learning gains, which are greater than 2. Therefore, a human teacher can be replaced by the hinting tutor without a significant decrease in learning gains. This conclusion is based on an interaction with the system or the teachers that lasted for 50 min, which is a usual duration for university course sessions and could be generalized in this context.

Applying the t-test again, there is no statistically significant difference in the initial knowledge level (pre-test grades) for both groups ($t = 0.000, n = 77, p = 1$), and the difference between the average students' initial knowledge level for both groups is within

Next, 4 hints are provided to you for choosing, you can choose all hints you want to
You will lose points for each hint you visualize

This is a hint about the first question, related to obtain the executable (You will lose 2 points for viewing the hint)

This is a hint about the second question, related to the clean target (You will lose 2 points for viewing the hint)

This is a hint about the third question, related to the substitution between lines 4-11 (You will lose 2 points for viewing the hint)

This is a hint about the fourth question, related to the substitution of line 2 (You will lose 2 points for viewing the hint)

Suppose the current directory contains the files main.c, part1.c, part2.c, makefile, clean where main.c uses the functions defined in part1.c and part2.c and the .c files can be compiled without errors. Select all the true statements related to this makefile, which code is provided next (each incorrect attempt subtracts 3 points)

Linea	Código
1	myprogram: main.o part1.o part2.o
2	gcc -o myprogram main.o part1.o part2.o
3	
4	part1.o: part1.c
5	gcc -c part1.c
6	
7	part2.o: part2.c
8	gcc -c part2.c
9	
10	main.o: main.c
11	gcc -c main.c
12	
13	clean:
14	rm -f myprogram main.o part1.o part2.o

☐ A single call of 'make myprogram' leads to the creation of an executable file ☐ A single call of make clean does produce errors ☐ Lines 4 to 11 can be substituted by a convenient implicit rule and it is equivalent ☐ Line 2 can be substitued by 'gcc -o $?' and it is equivalent
SCORE: 0.0/10.0

Fig. 2. Exercise with hints executing within XTutor (second phase of the interaction).

$[-0.77, 0.77]$ with a 90% probability. Therefore, the differences in students' initial knowledge levels for each group did not have a lot of influence on the results.

There was a low student-to-teacher ratio (13.5) in the context of typical university classroom sessions. With this ratio of 13.5, teachers were able to respond to most requests for help from students, as was confirmed by the interviews with teachers. Therefore, the availability of teachers for helping students during the experiment cannot be considered to be a key factor that influenced the results. Changing the student-to-teacher ratio, though, might influence the results.

Finally, an analysis was performed of the differences between the instruction given by human teachers and that given by the computer e-learning system during the experiment. These differences are what may affect the different learning gains. A human teacher uses some techniques of hint generation that are not used in the hinting module, but the hinting module uses some techniques that are not usually used by human teachers. From the results, these different factors might be compensating for each other in some way, reaching similar learning gains. Qualitative methods were used, interviewing the participating teachers, in order to understand their hinting techniques.

Features used by both teachers and the hinting tutor were the following.

- Different levels of hints. First, some initial hints are provided, but if the student is not able to solve the problem, then more detailed hints might be shown.

- Hints can be provided only after an incorrect response by the student, or they can be provided directly.
- Sequence of hints. A hint is composed of several ordered steps.

Techniques used by human teachers but not in the hint system were the following.

- Adaptive and personalized hints depending on the students. Teachers might know their students well and can thus assess which type of hint is better for each student depending on his/her lacks, attitudes, level, and so on. The hint specification allows the provision of adaptive hints in the XTutor system, but this feature was not enabled during the experiment.
- Hints are made available depending on the time students spent on solving the problem.

Finally, the following features were employed by the system tutor, but not usually by the teachers:

- hints can be in the form of text, or of other problems;
- scoring assignment for a problem and a penalty for each incorrect attempt;
- scoring penalty for viewing hints;
- scoring penalty for incorrect answers to hint problems;
- hints for which there is a maximum number that can be selected. There is some meta-information available to let students decide which hints to see.

In addition, other factors might have an important influence in the results, such as the shyness of some students, which can make them

TABLE I
GRADES OBTAINED BY BOTH EXPERIMENTAL GROUPS

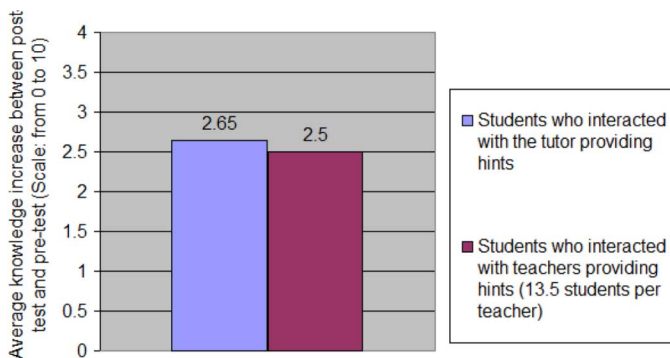| Grades Obtained	Mean	Std. Deviation	Std. Error
Pre-test for the group with human teachers	3.75	1.93	0.37
Post-test for the group with human teachers	6.25	1.67	0.31
Pre-test for the group with hinting tutor	3.75	1.95	0.27
Post-test for the group with hinting tutor	6.40	1.63	0.23

Fig. 3. Means of knowledge increase for both experimental groups.

reluctant to ask teachers but does not hinder their requesting hints from a computer-based system; or the quicker processing capability of a computer rather than a human, so a student can have faster access to hints without waiting for a teacher's answer; or the reduction of the monotony with a computer e-learning system.

V. CONCLUSION

This work has proved that both of the methods compared produced a significant enhancement of student learning during a classroom session of the course Computer Architecture Laboratory, namely a hinting computer e-learning system and human teachers generating hints.

The experiment did not show statistically significant results that the e-learning system is better than the teachers nor that the teachers are better than the e-learning system. However, they do suggest that human teachers can be replaced by the hinting e-learning system without a significant loss of effectiveness. The difference between the hinting tutor and human teachers would be in the interval [-0.38, 0.70] with a 95% probability, so in the worst case, this difference would not be significant with respect to absolute learning gains higher than 2.

A future experiment including adaptation capabilities for the hinting computer e-learning system is planned. With such adaptation capabilities, the learning gains obtained with the e-learning system might be greater, and thus better than human teachers.

REFERENCES

[1] R. DiPaolo, A. Graesser, D. Hacker, and H. White, "Hints in human and computer tutoring," in *The Design of Instruction and Evaluation: Affordances of Using Media and Technology*, M. Rabinowitz, F. C. Blumberg, and H. T. Everson, Eds. Evanston, IL: Routledge, 2004, pp. 155–182.

[2] K. VanLehn *et al.*, "Andes physics tutoring system: Five years of evaluations," in *Proc. 12th Int. Conf. Artif. Intell. Educ.*, Amsterdam, The Netherlands, 2005, pp. 678–685.

[3] E. Guzma and R. Conejo, "Self-assessment in a feasible, adaptive Web-based testing system," *IEEE Trans. Educ.*, vol. 48, pp. 688–695, 2005.

[4] V. Aleven, K. R. Koedinger, and K. Cross, "Tutoring answer explanation fosters learning with understanding," in *Proc. 9th Artif. Intell. Education Conf.*, Amsterdam, The Netherlands, 1999, pp. 499–506.

[5] K. N. Martin and I. Arroyo, "AgentX: Using reinforcement learning to improve the effectiveness of intelligent tutoring systems," in *Proc. 7th Int. Conf. Intell. Tutor. Syst.*, Maceió, 2004, pp. 564–572.

[6] P. J. Muñoz-Merino and C. D. Kloos, "A software player for providing hints in problem-based learning according to a new specification," *Comput. Appl. Eng. Educ.*, vol. 17, pp. 272–284, 2009.

[7] K. R. Koedinger, J. R. Anderson, W. H. Hadley, and M. A. Mark, "Intelligent tutoring goes to school in the big city," *Int, J. Artif. Intell. Educ.*, vol. 8, pp. 30–43, 1997.

[8] E. Harskamp and N. Ding, "Structured collaboration versus individual learning in solving physics problems," *Int. J. Sci. Educ.*, vol. 28, pp. 1669–1688, 2006.

[9] J. P. Henk, E. Harskamp, C. Suhre, and M. J. Goedhart, "How indirect supportive digital help during and after solving physics problems can improve problem-solving abilities," *Comput. Educ.*, vol. 53, pp. 34–50, 2009.

[10] C. M. Douglas, B. J. Reiser, M. Ranney, and J. G. Trafton, "Tutoring techniques: A comparison of human tutors and intelligent tutoring systems," *J. Learn. Sci.*, vol. 2, pp. 277–305, 1992.

[11] S. Suebnukarn and P. Haddawy, "A Bayesian approach to generating tutorial hints in a collaborative medical problem-based learning system," *Artif. Intell. Med.*, vol. 38, pp. 5–24, 2006.

[12] P. J. Muñoz-Merino, C. D. Delgado Kloos, M. Muñoz-Organero, and J. F. Naranjo, "Design and data analysis of exercises with hints," in *Proc. ASEE/IEEE FIE*, Saratoga, NY, 2008, pp. F2B15–F2B20.