Designing a Useful Problem Solving Coach: Usage and Usability Studies

Qing Ryan[*], Evan Frodermann[*], Kenneth Heller[*], Leonardo Hsu[†], Bijaya Aryal[¶]

[*]*School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455*
[†]*Department of Postsecondary Teaching and Learning, University of Minnesota, Minneapolis, MN 55455*
[¶]*Center for Learning Innovation, University of Minnesota, Rochester, MN 55904*

Abstract. Designing useful computer coaches for problem-solving in introductory physics requires an iterative process to develop both the software framework and content of the coaches. Research is necessary to determine which students use the coaches, how those students use the coaches, and whether the coaches are effective for those students. We report results of a study of prototype coaches to determine the user characteristics and usage characteristics. We also discuss how this data will guide the next iteration of these coaches.

INTRODUCTION

The ability to solve problems is highly valued and critically important to scientists and engineers [1]. Teaching problem solving based on the cognitive apprenticeship model [2] requires exhibiting all the requisite mental processes and practicing those processes while receiving real-time guidance and feedback (coaching). Providing this comprehensive coaching can be difficult because problem solving requires many decisions that students and even instructors do not recognize. Because students differ in their learning styles and rate of learning it is helpful if coaching is available at students' convenience. One approach to expanding students' access to effective coaching is the development of web-based computer coaches which could simulate the guidance provided by a moderately skilled instructor.

Designing a software framework for computer coaches is an iterative process. After small scale laboratory testing of computer coaches [3] we built a complete set of prototype coaches for introductory mechanics based on research in problem solving and computer tutoring. Here we report the results of testing these prototypes with students in large introductory physics classes to determine their potential usage and student perception of their utility. These results are being used to guide further development of both the software framework and the content of the coaches. Several cycles of implementation, assessment, and development will likely be necessary to achieve a useful and effective software framework.

Our goal is to determine if web-based computer coaches have a place in a multi-faceted toolkit to help diverse students develop problem-solving skills in an introductory physics class. So far, we have built and tested 35 coached physics problems spanning six major topics in an introductory mechanics class. The structure of these coaches is described elsewhere, along with the results of a pilot study in Fall 2011 [4].

In the Fall 2011 study, students could choose to do their homework either by submitting the correct answer to WebAssign (http://www.webassign.com/) within three tries or by completing the computer coach for the same problem. In that situation, students found the coaches very attractive, attempting (finishing at least the first section) an average of 28 (80%) and completing an average of 23 (66%). In Spring 2013, we conducted a second study designed to make use of the coaches less attractive to get a population of users and non-users in the same class. Below, we describe this study and discuss the questions (1) *What subpopulation of a class tends to use such coaches?* (2) *How did they use the coaches?* (3) *Do they perceive the coaches as useful?* We also discuss how this data will influence the development of the software framework.

THE STUDY

In Spring 2013, two sections of the calculus-based introductory mechanics course (249 students total) used the coaches. The results for the sections were similar so we have combined all of the data.

The two sections had different lecturers, but both focused on problem-solving facilitated by Cooperative Groups in the labs and discussion sections [5]. Students submitted weekly homework assignments

(10% of the course grade) through WebAssign and were allowed 5 tries to earn credit. Roughly one third of these problems were Context-Rich problems [6] on which students could get help from a computer coach. Students received no direct credit for using the coaches. The WebAssign and coached versions of a problem differed only in the symbols used to represent quantities in the problem.

Students took four midterm tests, each with two free-response Context-Rich problems, and a final exam with five Context-Rich problems. The five final exam problems were identical for both sections. In addition to the scores and written solutions to these problems, other collected data included pre and post test scores on the Force Concept Inventory (FCI) [7], a Math skills test, the Colorado Learning Attitudes about Science Survey (CLASS) [8], and a survey about the students' background. Students' use of the coaches was monitored by recording their keystrokes. They were also surveyed at the middle and the end of the semester about their use of the computer coaches.

USER CHARACTERISTICS

In any course, some students will tend to use resources such as the computer coaches and others find them unnecessary or incompatible with their personal preferences. To design effective coaches, one needs to know the relative sizes and characteristics of each group.

In Spring 2013, we were able to form three groups of students from the 249 total students (70% male (m), 30% female (f)) in the two sections: an L (light/non-user) group of 72 students (85% male, 15% female) who used 0–20% of the coaches[1], an M (medium-user) group of 38 students (55% m, 45% f) who used 40–60%, and an H (heavy-user) group of 49 students (65% m, 35% f) who used 80–100% of the coaches. We subtracted out people in the between 60-80% and 20-40% to get non-overlapping populations.

One difference between the three groups is in their gender. The proportion of females in the L group is about half that of the class as a whole. This is consistent with research that females with the same performance as males are more willing to seek assistance [9].

The three groups also differed in their physics preparation as measured by their scores on standardized assessments. Table 1 shows the pretest scores of the three groups, broken down by gender. The number (N) differs from those from the entire class because only students who took all three pre-tests

---

[1] Although there were 35 coaches, only 29 total coaches were considered for the data analysis because a database error made it impossible to track the usage of the first 6 coaches.

are included. A higher FCI pre-test score is correlated with lower use of the computer coaches. There is some indication that this may also be true for the Math skills

TABLE 1. Pre-test scores (as percentages) of the 3 groups.

Test	L (N=48)		M (N=27)		H (N=35)	
	m	f	m	f	m	f
Percent	85	15	67	33	66	34
FCI	58±5	59±12	53±7	42±7	46±5	31±4
MATH	58±5	66±8	53±6	61±9	54±15	45±4
CLASS	62±4	55±7	66±5	66±4	65±4	56±4

TABLE 2. Expectations of the 3 groups.

	N	Weekly study time (hrs)			Expected grade	
		≤5	6-10	≥10	A	B
L	48	25%	46%	29%	71%	29%
M	27	4%	59%	37%	70%	30%
H	35	8%	63%	29%	40%	60%

TABLE 3. Problem-solving score on the final exam (%).

Final Exam	m	f	Total
L (N=72)	69.2 ± 3.8	70.9 ± 9.8	69.5 ± 5.2
M (N=38)	64.8 ± 5.0	68.3 ± 5.8	66.4 ± 5.4
H (N=49)	72.8 ± 2.8	67.5 ± 5.8	71.0 ± 4.1

test. One might infer that the more poorly prepared students recognize this and choose to use easily accessible help.

Another difference among the groups is their expectation of the effort required for the class. Table 2 shows the results from a beginning-of-the-semester survey. Students in the L group expected to spend less time studying and to earn a higher grade in the class. No student expected to earn less than a B.

One might thus infer that students in the L group have high confidence in their ability to perform well. Students in the M group similarly expect to do well, but also expect to spend more time doing so. Students in the H group expect to spend more time and are less confident of their success.

Finally, we compare the performance of the three groups on the final exam. Table 3 shows averaged scores from 4 of the 5 free-response Context-Rich problems on the final exam (one problem that was graded anomalously in one section was dropped from the analysis). The scores for all three groups are not significantly different despite differences in FCI and MATH pre-test scores, while a combination of those scores usually accounts for about 25% of the difference in problem-solving performance on the final exam. This may indicate the coaches have helped the less-prepared students improve to the level of their classmates.

USAGE CHARACTERISTICS

In addition to characterizing the students using the

coaches, it is also important to examine how they use the coaches. There are two main differences in how the three groups of students used the coaches.

The first difference is the pattern of usage. Figure 1

FIGURE 1. Fraction of coaches associated with each test used by students in each group. The tests were given at the end of week 4, 7, 11, and 15. The lines are drawn to guide the eye.
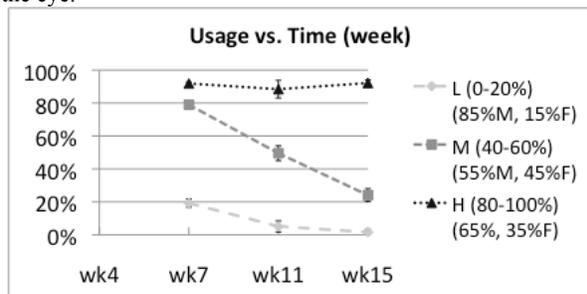
shows the fraction of the coaches preceding each midterm test used by each group. The L group used only 20% of the coaches associated with the topic of the second midterm before the second midterm. Their usage then dropped.[2] In contrast, students in the H group used the coaches consistently throughout. The M group started out using a high fraction of the coaches but their usage dropped steadily throughout the semester.

One possible reason is that the M students became more confident problem solvers, and believe that they no longer need the coaches. A second is that the M group decided that the coaches were no longer useful or valuable.

A second difference is how students reported using the coaches. On the end-of-semester survey, students were asked to select one of several choices describing how they used the coaches, or to write their own answer. The most popular choice was "I tried to solve the problems on my own and used the computer coaches for help if I got stuck," selected by 42% of the H group, 70% of the M group and 48% of the L group. The next most popular selection "I worked through the computer coaches before trying to solve the problems on my own," also differed depending on the group. While only 4% of the M group and 3% of the L group selected this option, 37% of students in the H group did so.

USABILITY & USEFULNESS

We gathered information about the coach's usability and perceived usefulness through the mid and end of semester survey, including rating statements on

a 5-point Likert scale, forced ranking on course components and free response questions.

TABLE 4. Students' responses to survey statements (%)

Statements	Strongly agree/Agree			Strongly disagree/Disagree		
	L	M	H	L	M	H
When using the computer coaches, it was usually clear how to proceed.	65	78	75	22	9	15
Using the computer coaches for homework made the homework take too long.	30	40	58	13	31	25
The computer coaches helped improve my conceptual knowledge of physics.	45	63	70	28	4	19
The computer coaches did NOT help improve my problem solving in this class.	34	11	21	41	74	67
Using the coaches improved my confidence when starting new, unknown problems.	47	70	70	26	4	16

TABLE 5. Usefulness ranking of class components

	L	M	H
Lectures	8.3 ± 0.3	7.5 ± 0.9	7.2 ± 0.7
Doing the homework	6.8 ± 0.5	7.2 ± 0.6	8.1 ± 0.4
Computer coaches	4.9 ± 0.5	7.2 ± 0.5	7.0 ±0.5
Tutor room	4.2 ±0.6	3.8 ± 0.8	4.3 ±0.6

The forced ranking question asked students to rank (with no ties) 10 different components of the class from most (10) to least (1) useful, both the H amd M groups ranked the coaches among the top 3 useful components (Table 5), ahead of other course components such as the textbook (5.40±0.76), labs (5.34±0.54), and problem-solving discussion sections (6.18±0.76).

From this data, we conclude that the interface of the coaches was reasonably clear and self-explanatory. Furthermore the target population believed that the coaches were beneficial to their learning at the end of the course. Indeed, the absolute FCI gain for the H (21%±5%) and M (19%±5%) groups was markedly higher than that of the L group (12%±5%), with no significant difference between gains for the male or female students. The keystroke data shows that the average time to complete a single problem using a coach was less than 31 minutes, comparable to the time spent by students interacting with a human coach in office hours. However, many students thought that the computer coaches took too long. 37% of the answers to the free-response question "What do you like least about the computer coaches?" mentioned that the computer coaches were either too long or too repetitive.

[2] Data on coaches used before the first midterm is not available because of a database error.

DISCUSSION

One of the goals of the current study was to assess students' perceptions of the utility of the computer coaches. In improving the usability of the coaches, one could choose to focus on improving the user experience and effectiveness for students who tend to use the coaches, or on trying to make the coaches more attractive to a larger fraction of students. Because most of the students in the physics course chose to use a substantial fraction of the coaches, the next iteration will focus on the former, which may, as a byproduct, lead to a larger user base.

Ideally, as students become more competent as well as confident problem solvers, one might expect to see a decrease in the use of the coaches. The fact that the H group not only continued to use almost all the coaches but also that a large fraction responded that they "…worked through the computer coaches before trying to solve the problems on my own," indicates that some mechanism is necessary to wean these students from the detailed help provided by the coaches. The new prototypes will be designed to better address the needs of the high and medium user population by having adjustable (by instructors or students) decision grain size. It will allow students to jump to sections of the problem solving framework that address their issues without repeatedly going through coaching they do not need. For the H group, it will allow them to have step-by-step coaching from the beginning to the end of a problem if desired. However, it will encourage bypassing detailed coaching. This flexibility might also reduce the perception of the coaches as rigid or repetitive. The additional flexibility should also engage students with different learning priorities at different times in the course. They would be able to use the same coach differently at the beginning and at the end of the course. This should be more useful to the M group whose usage decrease could be due to a perceived inefficiency of the coaches toward the end of the semester.

To produce the first set of prototype coaches to test with students, the decision structure and its software framework were basically procedural in nature. There were three different types of coaches with differing amounts of flexibility [3] and emphases but only one type of coach was available for each problem. The most popular and numerous type of coach (type1-based on students' survey response) had limited flexibility in that it would follow reasonable student choices in the problem solution but was rigid in the order of decisions that constituted those choices. Instructors wishing to modify the coaching pattern or build coaches for different problems needed knowledge of the underlying software language with more significant changes in procedure requiring more sophisticated software knowledge.

The second round of prototype coaches, now under construction, will address these issues of flexibility for the students and ease of modification for instructors. Indeed, the new software framework will allow instructors to build a new coach with no knowledge of the underlying software using only a graphical user interface (GUI). Although not part of the current study, this requirement was identified in a workshop on the coaches for physics instructors.

A computer coach is limited in its ability to help students identify their difficulties and remediate them and for this reason is not intended to replace a good human coach. However, when developed, computer coaches could provide a helpful approximation of the office hour experience that is available on demand and with whatever repetition desired by the student. We expect that most students will still need human intervention provided by the instructor and other students to make significant learning gains. Nevertheless, we hope that computer coaches interacting with students on the internet can be a flexible tool to support the learning of a diverse set of students in the introductory physics course.

REFERENCES

1. K. Schultz and J. Lochhead. In *Toward a unified theory of problem solving: Views from the content domains,* edited by M. U. Smith (Lawrence Erlbaum Associates, Hillsdale, NJ, 1991), pp. 99-114.
2. A. Collins, J. S. Brown and S. E. Newman, In *Knowing, learning and instruction*, edited by L. Resnick, (Lawrence Erlbaum, Hillsdale, NJ, 1989).
3. L. Hsu and K. Heller, *Proceedings of Physics Education Research Conference,* Sacramento, **790** (AIP, Melville, NY, 2004), pp. 197-200.
4. Q. Xu, K. Heller, L. Hsu and B. Aryal, *Proceedings of Physics Education Research Conference*, Philadelphia, **1513** (AIP, Melville, NY, 2012), p. 434-437.
5. P. Heller, R. Keith and S. Anderson, *Am. J. Phys.* **60**, 627-636 (1992).
6. P. Heller, and M. Hollabaugh, *Am. J. Phys.* **60**, 637-644 (1992).
7. D. Hestenes, M. Wells and G. Swackhamer, *Physics Teacher* **30**, 141-158 (1992).
8. W.K. Adams, K.K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein and C. E. Wieman, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010101 (2006).
9. M. Addis and J. Mahalik (2003). *American Psychologist*, **58**(1), 5-14.