

Computer problem-solving coaches for introductory physics: Design and usability studiesQing X. Ryan,^{1,*} Evan Frodermann,¹ Kenneth Heller,¹ Leonardo Hsu,² and Andrew Mason^{1,3}¹*School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA*²*Department of Postsecondary Teaching and Learning, University of Minnesota, Minneapolis, Minnesota 55455, USA*³*Department of Physics and Astronomy, University of Central Arkansas, Conway, Arkansas 72035, USA*

(Received 16 September 2015; published 16 February 2016)

The combination of modern computing power, the interactivity of web applications, and the flexibility of object-oriented programming may finally be sufficient to create computer coaches that can help students develop metacognitive problem-solving skills, an important competence in our rapidly changing technological society. However, no matter how effective such coaches might be, they will only be useful if they are attractive to students. We describe the design and testing of a set of web-based computer programs that act as personal coaches to students while they practice solving problems from introductory physics. The coaches are designed to supplement regular human instruction, giving students access to effective forms of practice outside class. We present results from large-scale usability tests of the computer coaches and discuss their implications for future versions of the coaches.

DOI: [10.1103/PhysRevPhysEducRes.12.010105](https://doi.org/10.1103/PhysRevPhysEducRes.12.010105)**I. INTRODUCTION**

Almost since the invention of the computer, educators and researchers in virtually every discipline have explored the roles that computers could play in enhancing instruction [1–3]. When evaluating such efforts, there are two important considerations, (i) will students use the system? and (ii) does the system help students? In this paper, we describe the development, implementation, and usability testing of a set of computer coaching programs designed to help students develop problem-solving skills in an introductory physics class. In a future paper, we will discuss the assessment of the educational impact of the coaches on students' problem-solving skills.

II. BACKGROUND

Problem solving is often cited as an essential skill for all citizens in a modern society [4] and it occupies a prominent position in the Framework for K-12 Science Education [5]. It is especially important for scientists and engineers. Problem solving by experts is an organized decision-making process with metacognition as a crucial component. It is often referred to as “nonroutine problem solving” to distinguish it from the learning of recipes that facilitate answering specific types of questions. This distinction was pointed out by Dewey [6] over a century ago and continues

to be made explicitly in national reports specifying educational needs for the 21st century [7]. Modern cognitive science defines problem solving as the process of reaching a goal when the path to that goal is uncertain [8]. This act of determining what to do when you don't know what to do is the type of problem solving necessary for a modern society and is what instructors desire their students to learn in an introductory physics course [9]. It is a recursive process that includes looping back through qualitative and quantitative analyses as well as recovering from dead ends. It is characterized by making judicious decisions within an organized framework. In contrast, questions that can be answered by the application of a known algorithm or procedure are often designated “exercises.”

Helping students develop real problem-solving facility is particularly appropriate for introductory physics, a gateway course for all STEM fields at the college level. Indeed, problem solving has long been seen as a way to facilitate students' construction of physics knowledge [10] and to familiarize students with the culture of science [11]. Nevertheless, although many introductory physics courses appear to emphasize problem solving, only a small fraction of students emerge from those courses with significantly improved problem-solving skills or an appreciation of what the process of problem-solving entails [12,13]. As evidenced by the large amount of research literature on this subject, this concern is not new [6,14,15] and continues to prompt calls for action [16].

Researchers have shown both in small-scale experiments and in classroom settings that it is possible, through targeted curriculum design efforts, to improve students' problem-solving skills [12,13,17]. However, one significant difficulty in implementing such efforts is that opportunities for students to practice solving problems while

*Corresponding author.
qxryan@cpp.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

receiving useful guidance and feedback are, at best, limited. Without coaching, students often practice using weak novice procedures rather than the expertlike frameworks they are taught [18].

One approach to increasing students' access to effective coaching is the development of computer coaches, software delivered via the Internet that can provide students with guidance and feedback. Similar to a human instructor engaging in a thoughtful Socratic dialog based on the needs and desires of a student, a computer coach could guide the student to the decisions that need to be made to construct a problem solution. A computer not only allows students to follow their own path to a solution but also to work at their own pace. Although not as flexible and insightful as a good human, computer coaches do have a number of advantages. A computer coach is available whenever a student desires coaching. They are infinitely patient and can be viewed as less judgmental than a human. Computer coaches cost very little to maintain once created and become more economical, while remaining equally effective, as they serve more students. Finally, computer coaches provide reproducible instruction that can be improved incrementally and systematically by input from the user community.

Numerous widely used online computer systems designed to grade students' answers to physics problems exist. For example, web-based homework systems, such as WebAssign (webassign.com), LON-CAPA (loncapa.msu.edu), smartPhysics (smartphysics.com), Expert TA (theexpertta.com), and Mastering Physics (masteringphysics.com) are easy for instructors to use and have large databases of problems drawn from and indexed to popular physics textbooks. These systems also provide various levels of help to the student in the form of hints, usually to correct students' mistakes. However, none of these systems coach students on the general decision-making skills that are critical to progressing toward expert problem solving.

Intelligent Tutoring Systems (ITSs) have also been developed for physics. Some of these efforts engage students in natural language dialogues [19,20] and introduce tools to help students solve quantitative problems [21,22]. Although such systems are more than a decade old, none are in wide use and their domains of applicability tend to be limited to one or two topics from introductory physics such as kinematics or Newton's laws, or a few concepts from electricity and magnetism. For example, in the 1990s, Reif and Scott at Carnegie Mellon University developed a modest computer coach called a Personal Assistant for Learning (PAL) [23,24]. Although some promising efficacy data with students were obtained, the system was never used appreciably outside of the institution at which it was developed, was limited to problems involving Newton's laws, and has now become technologically obsolete.

Probably the best known and most extensively developed ITS for physics is Andes [25]. Andes incorporates an artificial intelligence system that attempts to determine the user's mental state and offers appropriate guidance and feedback. Andes is designed to be a minimally invasive tutoring system, and thus does not provide coaching targeted to the general decision-making skills that are critical to problem solving. To the student, it can appear that Andes is focused on an equation-driven approach to problem solving. Despite positive assessment data [25] and the inclusion of users' guides to help students and instructors use the system, Andes is not widely used.

Thus, although there has been no lack of effort to design computer systems to help students solve problems in physics, existing systems have shortcomings that limit their usability. Next, we describe both our pedagogical and our technical approaches to constructing computer coaches that might overcome these drawbacks.

III. DESIGN OF COMPUTER COACHES

A. Pedagogical design

To help students move from a novice state of problem solving towards more expert practice, it is useful to examine the differences between novice and expert problem solvers. It is important to note that we do not expect students to become expert problem solvers within one or even a few semesters of physics [26]. The coaches are designed to help students move toward expertise by making competent decisions based on an expertlike framework [27,28].

Briefly, there are two major differences between expert and novice problem solvers: their knowledge organization and their problem-solving decision-making process (for general reviews of the literature see Refs. [12,13,29]). Experts organize their knowledge in interconnected chunks, hierarchically grouped around a small number of fundamental principles [30,31] and have organized decision-making processes that help them choose relevant principles for solving a problem [32]. In contrast, novices have fragmented or weakly connected knowledge, and their decision-making processes are often narrowly context related. Broadly speaking, a novice believes that each problem has a specific recipe of actions for solving it while an expert has a general decision-making process whose outcome is a set of actions that leads to a solution.

An expertlike problem-solving process is distinguished by the initial performance of a qualitative analysis to constrain the problem and categorize it based on fundamental principles [18]. Experts then apply their selected principle or principles to the problem in an organized manner, using self-monitoring strategies to assess their progress towards the solution [33]. Novices, on the other hand, typically focus on specific quantities in the problem and try to match those with mathematical procedures,

which they often call formulas. While novices might perform a rudimentary qualitative analysis when required to do so, they usually do not do so spontaneously [34] and often do not connect that analysis to the problem solution.

The process of moving toward expertise involves development in both these areas [35]. The organization of one's knowledge affects one's problem-solving process and vice versa. Indeed, improving students' generalized problem-solving skills can result in an improvement in their conceptual knowledge [36,37]. As described above, researchers and curriculum designers have shown repeatedly that it is possible, through targeted efforts, to improve students' problem-solving skills [12,13,17] and the common thread running through those efforts is that they are all explicitly or implicitly based on the cognitive apprenticeship.

Cognitive apprenticeship [38] is a theoretical framework that starts from the premise that human learning is complex and its details are specific to an individual in ways that could be unknown to the instructor. Despite this complexity, one type of pedagogy, apprenticeship, has been extremely successful throughout history and across cultures. In a cognitive apprenticeship, the functions of an apprenticeship are adapted to the context of formal education. This pedagogy can be found in most graduate education.

In a cognitive apprenticeship, the process of teaching incorporates the actions of modeling, coaching, and fading, all supported by temporary instructional tools called scaffolding. Essentially, modeling is showing students precisely what they need to do to accomplish an authentic task. A crucial part of modeling is to make all of the expert's intellectual processes of decision making visible to the student. Coaching is the process of giving students real-time feedback as they attempt a task by following, in their own way, their perception of the modeled process. Fading consists of allowing students to do the task themselves with reduced guidance and feedback. Scaffolding is temporary support, or "training wheels," that is removed as students become more proficient. All of these actions take place in what is called the environment of expert practice, where tasks include a meaningful context, motivation, and outcome [39].

The cognitive apprenticeship pedagogy is consistent with neural science that recognizes learning as the rewiring of neural connections. Learning becomes more meaningful as different neural networks are linked and the links between neurons are strengthened if they fire in close temporal proximity [40], providing a mechanism for learning being strengthened by practice and repetition. This biological picture supports the cognitive apprenticeship approach over the behaviorist quest for learning in pieces that then automatically become connected into a complex thought process such as problem solving.

Thus, in physics curricula designed to support the learning of problem-solving skills, the instructor models the use of an organized problem-solving framework similar to that expected of the student, taking care to make the decision-making processes involved visible and explicit. Interspersed with this modeling, the students receive coaching consistent with their own inclinations as they practice using this decision-making process to solve appropriate problems. Fading takes place when the amount and focus of the coaching changes as the students become more competent at solving problems, typically demonstrated in homework and tests. Scaffolding could include specific problem-solving frameworks and problems designed to encourage expertlike behavior and discourage novice behavior. It is important to note that, like problem solving, learning is a recursive process, and so instruction constantly cycles among these stages of modeling, coaching, and fading, rather than simply progressing from one to the next.

The computer coaches utilize two main instructional strategies, both of which are compatible with the cognitive apprenticeship framework. The first is a modified form of reciprocal teaching [41], originally developed to help middle-school students learn to read with good comprehension. To implement reciprocal teaching, we developed two types of coaches: type 1 (computer coaches student) and type 2 (student coaches computer).

In a type 1 coach, the computer models an organized decision-making framework to guide the student's problem-solving process. The student is asked to make the decisions necessary for solving a physics problem (e.g., choosing what to include in a picture or diagram, what physics principles to use, or how to apply those principles), with distractors based on known student difficulties. The computer gives feedback for each decision and requires that the student make a correct choice before moving on to the next decision in the process. Decisions can have more than one correct choice allowing the student to follow potentially fruitful, though not necessarily optimal, solution paths that appeal to them. A screenshot from a type 1 coach is shown in Fig. 1.

In a type 2 coach, the student and computer roles are reversed. The student chooses the decisions to be made by the computer. The computer makes those decisions, but may deliberately make mistakes corresponding to common student difficulties. The student must assess the computer's decisions and make any necessary corrections. Because the computer's responses are designed to reflect common student behavior, this coach also gives students practice in the important problem-solving process of debugging. The computer also acts in an oversight mode, assessing the student's responses and giving feedback. A screenshot from a type 2 coach is shown in Fig. 2. Within the cognitive apprenticeship framework, these two types of coaches provide students with step-by-step coaching through the solution process with different scaffolding.

Section One: Focus the Problem

Focus the Problem

- Picture
 - Important Objects ✓
 - Motion ✓
 - Position ✓
 - Velocity ✓
 - Acceleration ✓
 - Time ✓
 - Interactions ✓
 - ② Forces 1
 - Other Elements ⑥
- Questions
- Approach
- Physics Principle
- System
- Relevant Times
- Relevant Info
- Describe the Physics
- Plan the Solution
- Execute the Plan
- Evaluate the Solution
- Summary

Forces

What is the direction of the normal force on the puck by the ramp? ③

Select the correct answer from the list below. Use the scroll bar if necessary.

- ↑ ④
- ↗
-
- ↘
- ↓
- ↙
- ←
- ↖
- None/0

Problem

At the State Fair you see people trying to win a prize at a game booth. They are sliding a metal disk shaped like a puck up a wooden ramp so that it stops in a marked zone near the top of the ramp before sliding back down. You estimate that you can slide

Picture

$\vec{a}_i = \vec{a}$ $\vec{a}_r = \vec{a}$
 $\vec{v}_i = 8.0 \text{ ft/s}$ $\vec{v}_r = 0$
 $t_i = 0$ $t_r = ?$
 $W = 2.5 \text{ lbs}$

Assumptions

Air resistance can be neglected,
The size of the puck can be neglected (we can always measure from its front edge)

Feedback: No. The normal force exerted on an object by a surface points perpendicular to the surface and away from it. ⑤

FIG. 1. Screenshot from a type 1 coach (computer coaches student). The display shows a partially completed picture ①. The computer specifies a step in the framework ② and asks the student to decide on the direction of a force ③. The student’s decision ④ is incorrect, and the computer provides feedback ⑤. A red number to the right of each step ⑥ indicates the number of incorrect responses the student made for that decision, while a checkmark indicates that the step was performed correctly the first time.

Although both the type 1 and type 2 coaches progress through the entire problem-solving framework, students must learn to solve problems without the extensive scaffolding. A third type of coach, type 3 (student works independently, computer gives feedback), emphasizes the fading part of the cognitive apprenticeship paradigm by using the instructional strategy of learning from well-studied examples [42]. In this type of coach, the computer presents a problem to a student, who is asked to solve it independently of the coach and then enter an answer. The coach does not assume that a correct answer means that the student has a correct solution, but also asks follow-up questions to verify the correctness of the student’s work at important milestones in the solution process. A student who cannot complete a solution can choose to get help by selecting the part of the problem-solving framework believed to be the difficulty. The coach asks questions to determine if this is indeed the point of difficulty and guides the student through the necessary decision-making process using an interaction similar to that in the type 1 coaches.

After providing help on just that part, the coach asks the student to resume solving the problem independently. If the student’s answer to the problem or to one of the follow-up questions is incorrect, the computer gives appropriate feedback and the student can then choose where to get help. A screenshot from a type 3 coach is shown in Fig. 3.

Having two instructional strategies and three types of coaches is similar to the structure used in the design of Reif and Scott’s PAL tutors [24]. Operationally, the coaches are much like the “*Choose Your Own Adventure*” books [43] in the sense that the program operates like a flowchart (with loops) with responses determined by a student’s input. The full set of 35 coaches developed to address the topics found in the first semester of the calculus-based introductory physics course for engineering and physical science students at the University of Minnesota can be found on the University of Minnesota’s Physics Education Research Group webpage. Each coach helps a student solve a single physics problem using one of the three types of interactions described above.

FIG. 2. Screenshot from a type 2 coach (student coaches computer). The display shows a completed picture ①. The student, acting as a coach, has decided on a step for the computer, in its role as a student, to perform ②, but it is not an appropriate step at this point. The computer, in its oversight role, gives the student feedback ③.

The scaffolding provided by the coaches lies not only in the decision-making guidance, but also in the context-rich problems selected for the coaches. Context-rich problems are a type of problem that enables student learning of both physics concepts and problem solving [36,37]. Briefly, they are designed to (i) be straightforward to solve using expertlike strategies, but be difficult to solve using novice strategies, (ii) require students to make decisions on how to proceed with the solution, and (iii) have a context and motivation that appear authentic to students. Characteristic (iii), in particular, is important in the cognitive apprenticeship framework of learning within an expert environment. An example of a context-rich problem is given in Fig. 4.

B. Technical design

Our computer coaches have some of the features of what are sometimes called modern intelligent tutors, in that they are built around knowledge of student learning, expert behavior, and effective pedagogy. However, we purposely

do not use the term “tutors” because the coaches are designed to supplement existing classroom instruction, rather than provide standalone instruction. The building blocks of the computer coaches can be described by a modified Wenger model [44] where domain expertise, pedagogical expertise, and a model of student behavior are built into the system. However, the coaches have no independent intelligence and can be modified as new information becomes available only by the instructor.

The domain expertise includes both an analysis of the hierarchical knowledge structures of experts and a task analysis of the procedural knowledge for solving problems in physics, which is encapsulated in a decision-making framework (shown in Fig. 5) similar to those articulated by Refs. [6] and [45]. This domain expertise is also based on research comparing expert and novice problem solvers [18,32]. The explicit teaching of such a framework to help students organize their decision-making process has been shown to help them become better problem solvers in physics [46]. Although represented as a sequence, the

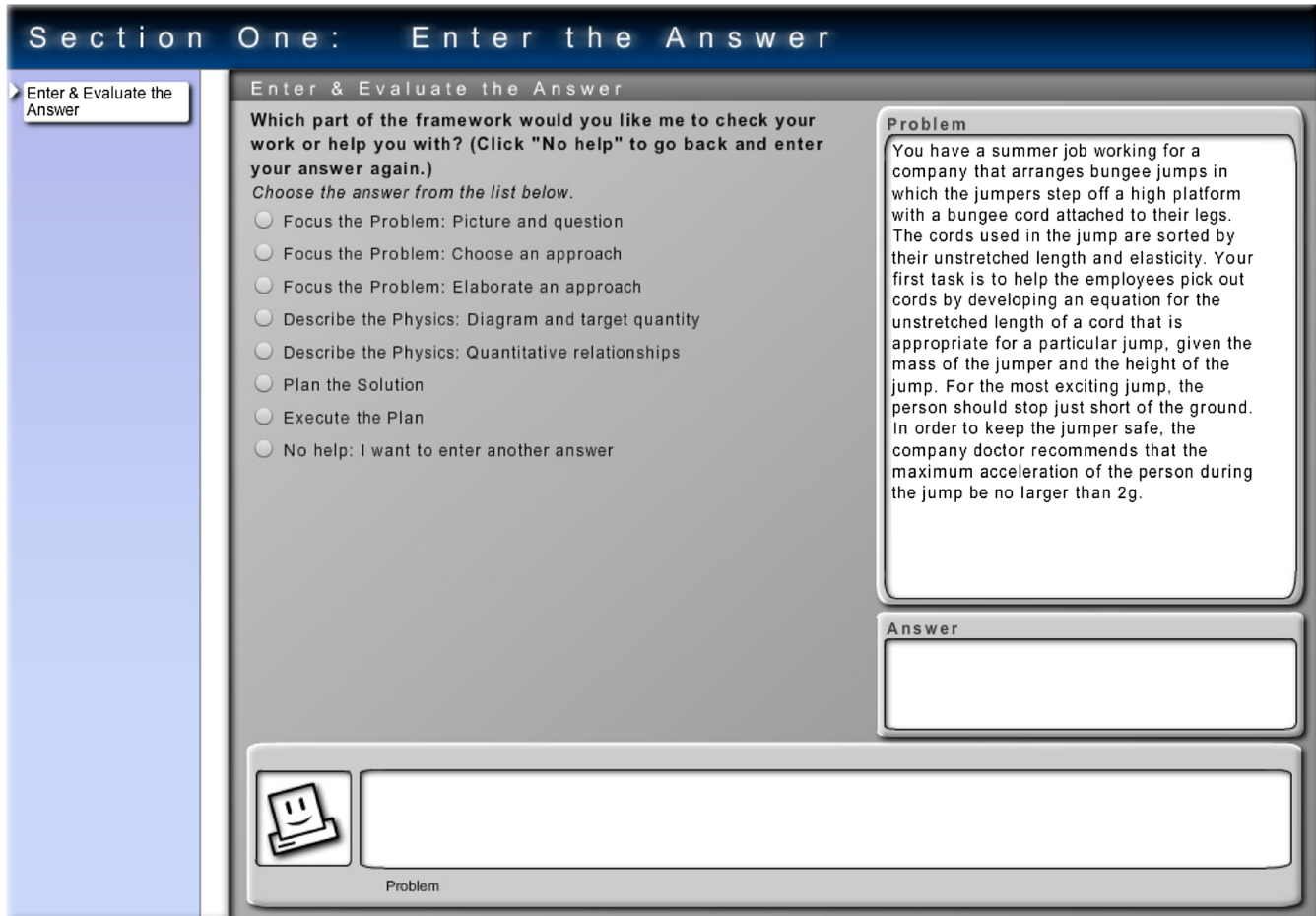


FIG. 3. Screenshot from a type 3 coach (student works independently, computer gives feedback). If the student gets stuck solving a problem or enters an incorrect answer, the computer asks the student to decide where in the problem-solving process the difficulty might occur and to get help.

solution path for a problem is recursive. A student typically begins with the first stage and proceeds to the second, but at some point, they will likely loop back to repeat a stage or skip forward past a stage.

The pedagogical expertise of the coaches arises from the research on effective pedagogies for teaching problem solving and physics instructors' pedagogical content knowledge. As described previously, the coaches rely on the well-established instructional strategies of reciprocal teaching and learning from well-studied examples in the context of cognitive apprenticeship.

At the train station, you notice a large horizontal spring at the end of the track where the train comes in. This is a safety device to stop the train so that it will not plow through the station if the engineer misjudges the stopping distance. While waiting, you wonder what would be the fastest train that the spring could stop at its full compression, 3.0 feet. To keep the passengers safe when the train stops, you assume a maximum stopping acceleration of $g/2$. You also guess that a train weighs half a million pounds. For purpose of getting an estimate, you decide to assume that all frictional forces are negligible.

FIG. 4. Example of a context-rich problem.

The model of student behavior is based on research on how novices solve problems [47,48]. Students tend to view solving a problem as knowing what to do instead of a process of finding out what to do by making a series of decisions. They often have difficulty visualizing a situation in sufficient detail to abstract meaningful information from that visualization, articulating the question posed by a problem embedded in a realistic context, using multiple representations of the situation, relating a problem to the fundamental principles of a field, making appropriate approximations needed to make a problem tractable, and determining whether their solution is likely to be correct. Because their knowledge and decision-making process are fragmented, they often do not know how to organize their ideas to initiate a problem solution, or even obtain useful help.

The primary need of a novice is to recognize problem solving as an organized decision-making process requiring metacognition. The necessary decisions include determining the relevance of their existing knowledge, connecting that existing knowledge to new knowledge about the situation, and determining any missing knowledge

necessary to arrive at a solution. For example, students often have difficulty determining what to include in a useful picture as part of a problem solution [49,50]. To build the initial student model, we have combined the existing literature with our own analyses of the physics problem-solving behavior of university physics students using their written solutions, videos of them solving problems individually and in groups, and interviews [37,50–53].

The coaches also take into account the guiding principles for the design of effective cognitive tutors [54]. For example, the coaches (i) communicate the goal structure underlying the problem by making the decision-making framework explicit, (ii) promote an abstract understanding of the problem-solving process by using the same explicit decision-making framework to solve all problems, (iii) minimize working memory load by maximizing the availability of relevant information in an easily accessible format, and (iv) provide immediate feedback on errors to reduce the amount of time students spend in unproductive mental states.

Similarly, the design of the coaches is consistent with well-known principles of multimedia learning [55,56]. They (i) use common web interactions such as clicking to select a statement or object, (ii) place two representations in close spatial proximity on the screen when translating from one representation to another, such as from a diagram to an equation, and (iii) pose questions using a conversational style. The student graphical user interface (GUI) was designed to be intuitive enough so that students could work through the coaches without any additional instruction either internal or external to the program.

The coaching programs themselves are written in Apache Software Foundation Flex (Flash), to provide a framework for student interaction, with World Wide Web Consortium XML to control the screen displays. An Oracle Corporation MySQL backend allows student responses to be stored on a database for subsequent analysis.

IV. USABILITY STUDIES

We conducted experiments to study three usability-related aspects of the computer coaches: (i) Are the coaches intuitive to students so that they are usable without any additional instruction or explanation? (ii) Do students perceive the coaches to be useful to their learning? and (iii) What are the characteristics of students who use the coaches more (or less) frequently? The first two questions are crucial because students must perceive any learning tool to be easy-to-use and beneficial to decide to use it. The third question is important because no pedagogical tool is attractive to and beneficial for every student.

A. Instructional setting

The studies described below were conducted in the first semester of an introductory calculus-based physics course

required for physical science and engineering majors at the University of Minnesota. The standard structure of this course, which includes three 50-min lectures (delivered in a room with auditorium-style seating), a 2-h laboratory section, and a 50-min discussion section each week, was maintained. During the lectures, the instructor modeled the use of the organized problem-solving framework shown in Fig. 5, as well as used informal group work such as Peer Instruction [57], partial problem solving by students, and interactive demonstrations. All of these techniques are typically used to some extent by all physics instructors at the University of Minnesota. Cooperative Group Problem Solving pedagogy [37] was used in the laboratories and discussion sections, which are taught in smaller sections of approximately 18 students by physics teaching assistants (TAs). The TAs are either physics graduate or advanced undergraduate students, assigned to sections based on scheduling considerations. All physics TAs receive a week of pedagogical introduction before the start of the semester and continuing support while they teach. During fall semesters, there are typically 5 lecture sections of this introductory course with about 200 students enrolled in each section, while in the spring, there are typically 2 lecture sections of the course with about 150 students enrolled in each section. Topics addressed in the course include kinematics, dynamics (Newton's laws and forces), conservation of energy, conservation of momentum, rotations, and oscillations. This has been the standard course design at the University of Minnesota for about 20 years and has already been shown to be successful at improving

Minnesota problem-solving framework

1. Focus the problem
 - Draw a picture, including given information
 - Determine question to be answered
 - Decide on useful physics principles to apply
2. Describe the physics
 - Draw diagrams and define physical quantities
 - Decide on the target quantity(ies)
 - Write down quantitative relationships
3. Plan the solution
 - Decide if/how quantitative relationships can be arranged to solve for the target quantity
4. Execute the plan
 - Use quantitative relationships to calculate target quantity
5. Evaluate the answer
 - Check if answer is properly stated
 - Check if answer is unreasonable
 - Check if answer is complete

FIG. 5. The Minnesota problem-solving framework [28].

TABLE I. The number of coaches for each topic and the types of coaches available for each topic.

	Kinematics	Dynamics	Energy	Momentum	Rotations	Oscillations	Overall
Type 1	2	2	4	3	2	1	14
Type 2	2	1	2	2	1	0	8
Type 3	2	3	2	3	2	1	13
Total	6	6	8	8	5	2	35

problem solving and conceptual knowledge in introductory physics [52]. The course typically has a D/F/W(withdraw) rate of approximately 5%.

The computer coaches used in these experiments consisted of 35 coached problems, each of which was available in one of the three types described previously. Table I shows the distribution of problem types for each of the topics addressed in the course. All 35 problems were context-rich problems. Many of the problems could be solved using more than one principle or combination of principles.

B. Experiment 1

During the Fall 2011 semester, the 35 computer coaches were made available to one of the five lecture sections, taught by one of us (L. H.), with 217 students who took the final exam. In that section, the homework for the entire course (worth 10% of the course grade) consisted only of the context-rich problems used in the 35 coaches. Students were allowed to satisfy their homework requirement either by submitting correct answers to the 35 problems through WebAssign (in three attempts or less), by completing the corresponding computer coach, or by some combination of the two methods. Each student's use of each coach was monitored by recording their keystrokes in a database. The WebAssign and coached versions of a problem differed only in the symbols used to represent quantities in the problem. During the semester, a database error prevented the complete logging of students' use of the six dynamics coaches, so the results presented here are based on the other 29.

We collected pre- and post-test scores on the Force Concept Inventory (FCI) [58], a math diagnostic test

(developed by the PER group at the University of Minnesota and available on our website [59]), and the Colorado Learning Attitudes about Science Survey (CLASS) [60]. In addition, during the first week of the course, students completed an 18-question survey regarding their background and expectations for the course (developed by the PER group at the University of Minnesota and available on our website). At the end of the semester, we collected students' solutions to the five free-response problems on the final exam and gave students a survey regarding their opinions about the computer coaches. There were two versions of the end-of-semester survey and students chose which to complete based on whether they thought they principally used WebAssign or the coaches to satisfy their homework requirement. The survey was delivered via WebAssign. Students received extra credit worth 0.5% of their final grade for completing this survey and their responses were not anonymous.

The coaches proved to be extremely popular with students. When given a choice between solving problems independently and submitting an answer using WebAssign or using the coaches to fulfill a homework assignment, an overwhelming proportion of students chose to use the coaches. Out of the 29 coaches for which there are complete records, students completed an average of 19 coaches. Only 28 of the 217 students who took the final exam completed fewer than 10 coaches. Students also used the coaches for help solving a problem without using them to get credit. Students in the class attempted an average of 22.5 of the 29 coaches and only 10 students attempted fewer than 10 of the coaches. To be counted as an attempt, students must have completed at least the first section of the coach, corresponding to the first stage of the Minnesota problem-solving framework shown in Fig. 5. When using a

TABLE II. Student usage of the coaches by course topic in fall 2011. *K* is kinematics, *D* is dynamics, *E* is energy, *M* is momentum, *R* is rotations including statics, and *O* is oscillations. The overall figures exclude the dynamics coaches. The data come from the student keystrokes recorded in the database.

	<i>K</i>	<i>D</i>	<i>E</i>	<i>M</i>	<i>R</i>	<i>O</i>	Overall
Completion	79%	N/A ^b	57%	54%	82%	61%	65%
Attempt	87%	N/A ^b	75%	70%	85%	71%	77%
Avg. completion time (min) ^a	35	27 ^c	23	28	35	30	29

^aSince the type 3 coaches ask students to solve the problem on their own first, average completion time was calculated from type 1 and type 2 coaches only.

^bA database error prevented the accurate computation of these numbers.

^cComputed based on incomplete data.

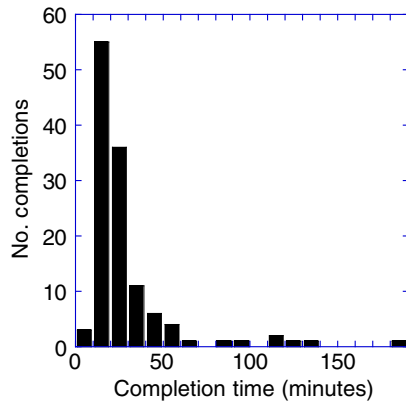


FIG. 6. Histogram of completion times for one particular type 1 coach. The average completion time was calculated using only the times that were within the main distribution (less than 70 min.).

coach without completing it, it was most common for students to complete the parts that helped with the first two stages of the Minnesota problem-solving framework, corresponding to a qualitative analysis of the problem and missing only the mathematics required to obtain an answer as well as to check it. Table II shows this usage data broken out by topic. The completion (attempt) percentage for a given topic is the fraction of total possible coach completions (attempts) for that topic. For example, for kinematics, out of 1302 possible coach completions or attempts (6 coaches multiplied by 217 students), 1026 coaches were completed and 1127 coaches were attempted (including completions).

Table II also shows that the average time taken for students to complete a type 1 or type 2 computer coach was just under 30 min. Type 3 coaches are not included in this calculation because students using a type 3 coach must independently try to work through a problem before engaging with the coach and the time a student spent working independently could not be recorded. The overall average is the mean of the average completion times for all type 1 and 2 coaches for which full records exist (not

including those for dynamics). When students spent more than an hour to complete a coach, the keystroke logs showed that the students took either one or more small breaks of between 10 and 30 min or a long break of more than 1 h. When computing the average completion time for a particular coach, we used only those times within the main distribution (see Fig. 6), eliminating the long tail of outliers that included those taking significant breaks. Taking long breaks was rare. Only 5% of the students took a break of more than 1 h.

Because students were allowed only 3 attempts to enter a correct answer using WebAssign, if those attempts were exhausted the only way for a student to get credit for a homework problem was to complete the computer coach. However, an analysis of the time stamps for both WebAssign and the computer coaches shows this occurred for at most 8% of the 6293 homework problems (217 students multiplied by 29 homework problems). Table III shows the time ordering of the use of the two systems for completing the homework problems. Only a small percentage of the problems (6% of 6293 homework problems) were not attempted at all (the “No effort” row). The largest noncompletion of homework (18% of 434 problems) was for the oscillations topic, which consisted of only two problems and was due during the last week of class. The three most popular methods used by students to complete the homework were to use only the coaches (44% of the 6293 homework problems), followed by using the coaches and then WebAssign (19%), and then using WebAssign only (17%). On average, students used the coaches either to completion or to help enough to solve the problem 70% of the time.

Students’ responses to the end-of-semester survey were consistent with their extensive use of the coaches. The survey consisted of 29 questions, which were a mix of free response, forced ranking, and multiple choice. This questionnaire was completed by 61% of the 217 students completing the course. Because of the high coach usage, no attempt was made to categorize the results by coach usage. Table IV shows student responses to four 5-point

TABLE III. The percentage of homework problems in a given topic completed using a particular time order of using WebAssign and the computer coaches. Headings are defined in Table II. The overall figures exclude the dynamics coaches. The data come from the time stamps recorded by both the coach and WebAssign. The sample size is computed by multiplying the number of different homework problems for a topic by the number of students in the analysis (217).

	<i>K</i>	<i>D</i>	<i>E</i>	<i>M</i>	<i>R</i>	<i>O</i>	Overall
Sample size	1302	1302	1736	1736	1085	434	6293
WebAssign only	9%	N/A ^a	21%	25%	8%	12%	17%
WebAssign first, then coach	10%	N/A ^a	7%	9%	8%	4%	8%
Interleaved use of both systems	11%	N/A ^a	6%	6%	6%	2%	7%
Coach first, then WebAssign	19%	N/A ^a	25%	18%	13%	15%	19%
Coach only	46%	N/A ^a	38%	36%	59%	49%	44%
No effort	5%	N/A ^a	4%	5%	7%	18%	6%

^aA database error prevented the accurate computation of these numbers.

TABLE IV. Student responses to Fall 2011 end-of-semester survey questions about the usability and utility of the computer coaches.

Survey statement	Strongly agree or agree	Neither agree nor disagree	Strongly disagree or disagree
When using the coaches, it was usually clear how to proceed.	90 ± 3%	8 ± 2%	2 ± 1%
The computer coaches did not help improve my problem solving.	8 ± 2%	13 ± 3%	80 ± 3%
The computer coaches helped my conceptual knowledge of physics.	81 ± 3%	14 ± 3%	5 ± 2%
The computer coaches helped me identify what I needed to get help with from other sources.	74 ± 3%	21 ± 4%	5 ± 2%

Likert scale questions about the perceived utility and usability of the coaches. In the Table, we have aggregated the Agree and Strongly agree responses, as well as the Disagree and Strongly disagree responses. The ranges are calculated using the standard error of the proportion. Student responses show that they thought that the computer coaches were easy to use and useful to their learning. Table V shows students' opinions on which type of computer coach was most useful to them at the beginning and the end of the course. At the beginning of the course students perceived the type 1 coaches to be the most useful while at the end of the semester all types were considered to be equally useful.

Finally, Table VI shows the results from a survey question asking students to rank 18 components of the course in order of perceived usefulness to their learning. Students were asked to rank a component only if they had used it. The computer coaches were essentially tied with lectures as the component that students perceived to be the most helpful to their learning, even higher than the human help available to them from physics TAs in a departmental tutor room or the instructor's office hours. They were also perceived to be more helpful than the peer and TA coaching they received in the Cooperative Group Problem Solving discussion sections that were also perceived as one of the most helpful elements of the course.

In summary, virtually all of the students used the coaches to some extent (only two of the 217 students completing the course did not complete any of the coaches and only one of those two did not attempt any of the coaches). Based on the students' survey responses, we conclude that most students found the coaches to be easy-to-use and that the interface was clear and self-explanatory. The type I coaches were judged to be the most useful at the beginning of the course

TABLE V. Fall 2011 student responses to a survey question asking them which type of coach they found to be the most or least useful at the beginning and end of the course.

	Type 1	Type 2	Type 3
Which type of computer coach did you find the most (least) useful at the beginning of the course?	63 ± 4% (5 ± 2%)	23 ± 4% (24 ± 4%)	14 ± 3% (71 ± 4%)
Which type of computer coach did you find the most (least) useful at the end of the course?	29 ± 4% (23 ± 4%)	33 ± 4% (29 ± 4%)	38 ± 4% (48 ± 4%)

while at the end, all three types of coaches were judged equally useful. Overall, the students perceived the coaches as being among the most useful elements of the course in terms of their learning, improving both their conceptual understanding and problem-solving skills in physics. Because such a large fraction of the class used the coaches, it was not possible to find a large enough sample to compare the characteristics of students who used the coaches to those that did not.

TABLE VI. Results from a fall 2011 survey question asking students to rank (with no ties) the usefulness of 18 components of the class to their learning. Students were asked to rank a component only if they had used it. Lower numbers are better.

Component	Percentage of respondents ranking this component	Average ranking (lower is better)
Lectures	100%	3.8 ± 0.3
Computer coaches	99%	3.9 ± 0.2
Practice quizzes	96%	5.1 ± 0.3
Discussion sections	100%	6.0 ± 0.3
In-class clicker questions	98%	6.1 ± 0.3
Studying with other students outside class	74%	7.3 ± 0.5
Quizzes	99%	7.9 ± 0.4
Student's own lecture notes	91%	8.4 ± 0.4
Posted lecture notes	73%	8.6 ± 0.4
WebAssign	87%	9.2 ± 0.4
Lab sessions	97%	9.3 ± 0.4
Textbook readings	88%	9.9 ± 0.4
Discussion with professor outside class	47%	10.2 ± 0.7
Writing lab reports	92%	10.4 ± 0.4
Textbook problems	78%	11.2 ± 0.4
TA help in tutor room	48%	11.2 ± 0.7
Supplementary text on problem solving	44%	11.8 ± 0.7
Discussion with TA outside class	48%	12.6 ± 0.7

C. Experiment 2

A second experiment, where using the coaches was made less attractive, was run during the spring 2013 semester. In this study the coaches were made available in both lecture sections of the introductory calculus-based mechanics class, one with 142 students and the other with 94 students who completed the course. The two sections were taught by two of the authors (E. F. and L. H.). The homework for the two sections was similar, but not identical, with the 35 coached problems making up about 30% of the total number of homework problems, but worth about 40% of the total homework credit. Homework problems that were not coached problems were selected from the end-of-chapter problems found in the textbook [61]. Unlike the fall 2011 semester, students received no homework credit for using the coaches. Students were required to submit their homework (worth 10% of the course grade) through WebAssign and were allowed 5 tries to enter the correct answer to receive credit. As in fall 2011, the WebAssign and coached versions of a problem differed only in the symbols used to represent quantities in the problem.

The two instructors consulted with each other on a regular basis to keep the two sections as parallel as possible. Both followed the same schedule of topics and exams, had identical laboratories and final exams, and gave midterms with isomorphic problems. Both instructors performed the same kinds of activities during lecture (as described previously in Sec. IV A). Because of these similarities between the two sections, they were combined for analysis purposes. Although the nature of our research questions (assessing the usability of the coaches) makes it unnecessary for the students in the two sections to be comparable, we found that they were ($51 \pm 2\%$ vs $50 \pm 2\%$ for FCI pretest scores, $61 \pm 2\%$ vs $60 \pm 2\%$ on a Math diagnostic test, and $63 \pm 2\%$ vs $65 \pm 2\%$ on the CLASS).

Information collected from students included the same data collected in fall 2011, the only difference being that students were surveyed on their opinions of the computer coaches twice, once at the midpoint of the semester (week 8) and once at the end, and only one form of the survey was given to all students, regardless of their coach use. Students received extra credit worth up to 1.4% of their final grade for completing both surveys and their responses were not anonymous. During the semester, a database error prevented the identification of a fraction of the students using the first six coaches dealing with kinematics, so the results presented here are based on the other 29.

In contrast to the fall 2011 semester, where most students used most of the coaches, the students in the spring 2013 semester showed a wide range of use, with students attempting an average of 12.8 of the coaches. This variation, when there was no direct incentive to use the coaches, allowed us to divide the students into groups based on the frequency of coach usage. For analysis purposes we defined the following user groups: a low-user

(*L*) group using between 0 and 25% (0 to 7) of the coaches, a medium-user (*M*) group using between 35% and 65% (11 and 18) of the coaches, and a high-user (*H*) group using between 75% and 100% (22 and 29) of the coaches. We include a gap of 10% between each user group to exclude intermediate cases.

Of the 236 students in the two sections that completed the course, 201 fell into one of the three user groups. The *L* group (17 females and 72 males) was 38% of the class, the *M* group (22 females and 32 males) was 23% of the class, and the *H* group (22 females and 36 males) was 25% of the class. The other 14% of the class were in the excluded boundaries between the groups. One observation is that there is a difference in the gender ratios of each of the user groups. While only 19% (17 out of 89) of the students in the *L* group were female, the *M* and *H* groups were 41% (22 out of 54) and 38% (22 out of 58) female, respectively. The percentage of female students in the class as a whole was 31% (72 out of 236).

We hypothesized that the three groups of students might differ in terms of their self-confidence and preparation at the beginning of the semester. To test this hypothesis, we examined students' conceptual preparation based on the FCI pretest, as well as their self-confidence based on two of the 18 questions on the precourse survey of the students' background and expectations for the course. These two questions asked students what grade they expected to get in the class and how many hours per week they expected to spend studying for the class. Since the students took the FCI concurrently with the survey, they did not know their FCI score at that time. Of the 89 students in the *L* group, 64 (12 females and 52 males) completed both the FCI pretest and background survey, while 42 (12 females and 30 males) of the 54 *M* students and 45 (18 females and 27 males) of the 58 *H* group students completed both. In all cases except for the female *M* students, over 70% of the students in each of the six subgroups completed both surveys, so the results should be representative of those groups (55% of the female *M* students completed both surveys).

Table VII shows the results from the two questions from the background survey as well as the FCI pretest. Because FCI performance is subject to a well-known gender effect [62,63] that we have verified in our students from past classes [64], FCI scores are broken out by gender, while responses to the survey questions are not. As can be seen, students in the *L* group differed from students in the *H* group in that *L* group students expected to receive a higher grade ($\chi^2 = 7.1$, $p < 0.01$) while spending less time studying for the course ($\chi^2 = 5.0$, $p < 0.05$). We interpret this as a difference in student confidence of their preparation for the course. Furthermore, the *L* group had a higher average pre-FCI score than the *H* group ($p < 0.005$ using a Kolmogorov-Smirnov test) indicating that the *H* users were, on average, less well prepared conceptually.

TABLE VII. Characteristics of students in the low (*L*), medium (*M*), and high (*H*) coach use groups from a background and expectations survey and the FCI administered during the first week of the spring 2013 semester.

Coach Use	Gender	Expected weekly study time (hrs)			Expected grade		FCI pre-test
		≤5	6-10	>11	A	B/C	
<i>L</i> (<i>N</i> = 64)	12 <i>F</i>	23 ± 5%	49 ± 6%	28 ± 6%	69 ± 6%	31 ± 6%	55 ± 7% (<i>F</i>)
	52 <i>M</i>						57 ± 2% (<i>M</i>)
<i>M</i> (<i>N</i> = 42)	12 <i>F</i>	12 ± 5%	62 ± 7%	26 ± 7%	71 ± 7%	29 ± 7%	39 ± 6% (<i>F</i>)
	30 <i>M</i>						55 ± 3% (<i>M</i>)
<i>H</i> (<i>N</i> = 45)	18 <i>F</i>	7 ± 4%	64 ± 7%	29 ± 7%	47 ± 7%	53 ± 7%	39 ± 4% (<i>F</i>)
	27 <i>M</i>						45 ± 3% (<i>M</i>)

We also examined the usage patterns of the three groups of students as a function of time. Figure 7 shows the fraction of coached problems attempted and the class topic during that time period. This fraction was calculated in the same way as the fractions in the analysis of experiment 1. The uncertainties again represent the standard error of the proportion. The database error during the kinematics portion of the course affected principally the problems assigned during week 3, so those data are not shown. The effect of the database error on the week 4 problems was small enough that its effect could be taken into account using larger error bars.

Three distinct patterns of usage can be seen. The students in the *L* group began using very few of the coaches, dropping to essentially none by the end of the course. On the other hand, the *H* group used over 80% of the coaches on average, remaining essentially constant throughout the entire semester. Students in the *M* group, however, showed a dramatic shift in coach use in the final third of the semester, using about 70% of the coaches for the first half of the homework assignments, and suddenly dropping to about 30% at the end of the semester.

A closer look at the exact timing of the change in usage for the *M* group finds that it is in close proximity to spring break but occurs one week after the students return. It also occurs at the midpoint between two tests and in the middle of a single homework assignment that included both conservation of energy and conservation of momentum problems. Within that single assignment, students in this group used a much smaller fraction of the coaches for the momentum problems (25%) than the coaches for the energy problems (81%).

Results from questions similar to those from fall 2011 on the spring 2013 mid- and end-of-semester surveys are shown in Table VIII. On the midsemester survey, 54% (48 out of 89) of the *L* students responded to the survey, along with 78% (42 out of the 54) of the *M* students and 98% (58 out of the 59) of the *H* students. On the end-of-semester survey, 90% (71 out of the 89) of the *L* students responded, as well as 83% (45 out of the 54) of the *M* students and 93% (55 of the 59) of the *H* students.

A clear majority of students in the *M* and *H* groups thought that the coaches were helpful to their learning of

physics and problem solving. Surprisingly, at the end of the course a plurality of the students (43 ± 6% conceptual, 45 ± 6% problem solving) in the *L* group also agreed, although, not as large a fraction as in the *M* (60 ± 7% conceptual, 77 ± 6% problem solving) and *H* (71 ± 6% conceptual, 71 ± 6% problem solving) groups. Furthermore, the fraction of students who were positive about the helpfulness of the coach increased from the middle to the end of the semester in most categories. Likewise, all groups thought the coaches helped them identify their difficulties and gave them confidence in solving unknown problems. In particular, this is true even of the *M* students, whose responses do not seem to reflect the dramatic drop in usage.

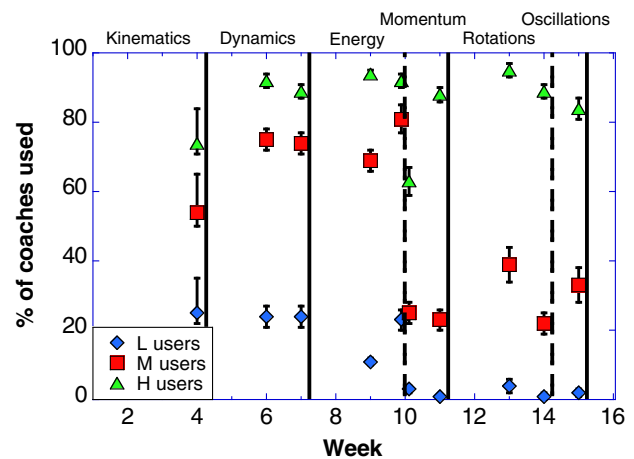


FIG. 7. Use of the computer coaches as a function of time throughout the spring 2013 course by students in each of the three usage groups. A database error prevents the calculation of student usage of the coaches in week 3, as well as causes the asymmetric error bars for the week 4 data. The solid vertical lines show the timing of the four in-class tests, which occurred at the end of weeks 4, 7, 11, and 15 of the class. The solid and dashed lines show the boundaries between different topics (labeled at the top of the graph). Usage by students in the *L* and *H* groups is relatively constant while usage by *M* students changes dramatically between the energy and momentum sections of the course, in the middle of the week 10 homework. Spring break occurs between weeks 8 and 9 of the course.

TABLE VIII. Spring 2013 student responses to questions from an end-of-semester survey about the usefulness of the computer coaches. Numbers in parentheses are student responses from a midsemester survey.

Survey statement	User group	Strongly agree or agree	Neither agree nor disagree	Strongly disagree or disagree
The computer coaches helped improve my conceptual knowledge of physics.	<i>L</i>	43 ± 6% (38 ± 7%)	30 ± 6% (35 ± 7%)	27 ± 5% (27 ± 6%)
	<i>M</i>	60 ± 7% (48 ± 8%)	31 ± 8% (26 ± 7%)	9 ± 4% (26 ± 7%)
	<i>H</i>	71 ± 6% (65 ± 6%)	13 ± 5% (16 ± 5%)	16 ± 5% (19 ± 5%)
The computer coaches did not help improve my problem solving in this class.	<i>L</i>	33 ± 6% (25 ± 6%)	22 ± 5% (27 ± 7%)	45 ± 6% (48 ± 7%)
	<i>M</i>	7 ± 4% (14 ± 5%)	16 ± 6% (19 ± 6%)	77 ± 6% (67 ± 7%)
	<i>H</i>	20 ± 5% (12 ± 4%)	9 ± 4% (11 ± 4%)	71 ± 6% (77 ± 6%)
The computer coaches did not help me identify what I was confused about.	<i>L</i>	26 ± 5% (34 ± 7%)	23 ± 5% (21 ± 6%)	51 ± 6% (45 ± 7%)
	<i>M</i>	16 ± 5% (24 ± 7%)	18 ± 6% (24 ± 7%)	67 ± 7% (52 ± 8%)
	<i>H</i>	15 ± 5% (23 ± 6%)	13 ± 5% (25 ± 6%)	72 ± 6% (53 ± 7%)
Using the coaches improved my confidence when starting new, unknown problems (e.g., quiz problems).	<i>L</i>	50 ± 6% (43 ± 7%)	27 ± 6% (33 ± 7%)	23 ± 5% (24 ± 6%)
	<i>M</i>	69 ± 7% (55 ± 8%)	24 ± 7% (29 ± 8%)	7 ± 4% (17 ± 6%)
	<i>H</i>	74 ± 6% (65 ± 6%)	11 ± 5% (21 ± 6%)	15 ± 5% (14 ± 5%)

Table IX shows results from the spring 2013 end-of-semester survey in which students were asked to rank 10 components of the course in order of perceived usefulness to their learning. As in fall 2011, students were asked to rank each component that they used with a number between 1 and 10 with no ties. The N 's are different for the three groups on this question than for the survey questions in Table VIII because only results from students who carried out the ranking procedure correctly as described in the instructions were included. As might be expected, the *L* students ranked the computer coaches significantly lower (7th out of 10) than the *M* and *H* students (3rd and 2nd, respectively). On an identical question on the midsemester survey, the *L*, *M*, and *H* groups ranked the computer coaches 8th, 4th, and 4th, respectively.

Table X shows results of a question in which students were asked to consider ten ways in which the coaches could have helped them, and to rank them from the item with which the coaches helped them the most to that with which the coaches helped them the least (without any ties). The top five choices were the same for all three groups: getting started solving a problem, interpreting the problem text, deciding what physics to use, applying the physics concepts to a specific problem, and applying the appropriate equations to a particular problem.

In summary, when the coaches were made available to students as source of help without any direct inducement,

we found a much wider range of use. Students who chose to use a larger fraction of the coaches (at least 75%) seemed to have characteristics associated with lower preparation and self-confidence for physics. All students were positive about the helpfulness of the coaches, even those who did not use many of them (no more than 25%). Of those who used a significant fraction of the coaches (at least 35%), they were rated among the most useful components of the course.

V. DISCUSSION

Our goal in constructing the coaches and testing them in introductory physics classes was to determine the feasibility of using computers to provide students with coaching in the decision-making process critical for solving physics problems.

From our results, we believe it is clear that such coaches can be constructed and be perceived as both usable and useful by students without any additional instruction or explanation. As can be seen from Table IV, about 90% of the students using the coaches in fall 2011 agreed or strongly agreed with the statement "When using the coaches, it was usually clear how to proceed." Furthermore, in fall 2011, when given the choice between using the coaches or a typical web-based homework system to complete their homework in the course, students

TABLE IX. Results from a spring 2013 end-of-semester survey question asking students to rank (with no ties), the usefulness of 10 components of the class to their learning, broken down by user group. Lower numbers are better.

<i>L</i> group (<i>N</i> = 58) (light users)	<i>M</i> group (<i>N</i> = 35) (medium users)	<i>H</i> group (<i>N</i> = 46) (high users)
Lectures 2.8 ± 0.4	Lectures 3.3 ± 0.7	Doing the homework 2.9 ± 0.4
Discussion section 4.0 ± 0.5	Doing the homework 3.6 ± 0.4	Computer coaches 3.7 ± 0.5
Doing the homework 4.4 ± 0.5	Computer coaches 3.7 ± 0.5	Lectures 3.9 ± 0.6
Clicker questions 4.8 ± 0.4	Discussion section 4.5 ± 0.7	Discussion section 4.4 ± 0.5
Labs 5.0 ± 0.5	Clicker questions 4.6 ± 0.6	Clicker questions 4.9 ± 0.5
Textbook 5.8 ± 0.5	Labs 5.4 ± 0.4	Textbook 5.6 ± 0.6
Computer coaches 6.0 ± 0.4	Textbook 5.8 ± 0.7	Labs 5.9 ± 0.4
Tutor room 6.6 ± 0.6	Suppl. problem-solving text 6.7 ± 0.8	Tutor room 6.7 ± 0.8
Feedback from WebAssign 6.8 ± 0.5	Feedback from WebAssign 7.4 ± 0.6	Suppl. problem-solving text 6.9 ± 0.6
Suppl. problem-solving text 7.0 ± 0.5	Tutor room 7.6 ± 0.8	Feedback from WebAssign 7.6 ± 0.5

overwhelmingly chose to use the coaches. Keystroke data showed that the average length of time for a student to complete one of the coaches was roughly 30 min, comparable to the amount of time a human instructor might spend with a student on a similar problem. Furthermore, students, on the whole, seemed to stay on task while working through a coach, without taking many breaks.

In spring 2013, when there was no direct credit for using the coaches, roughly 50% of the students, those not belonging to the *L* user group, used more than 70% of the coaches to get help with their homework during the first half of the course. This number decreased by the end of the course primarily due to the drop-off in use by about 1/4 of the class (the *M* users) during week 10. When ranking the coaches as a component of the class useful to their learning, the coaches ranked among the top 3 most useful for those that used at least 35% of them. All students, even those that used less than 25% of the coaches, ranked them higher than

other out of class individual help such as the tutor room staffed by TAs, a problem-solving book [28], and feedback from the electronic homework system.

Survey data show that a majority of the students in the class thought that the coaches were useful for helping them improve their conceptual knowledge of physics, problem-solving skills, and their confidence in solving new, unknown problems. In particular, the relatively high opinion of the coaches held by *M* students at the end of the semester suggests that the reason they stopped using the coaches mid semester was not because they did not value them.

Students believe that the coaches help them not only with their problem solving, but also with their conceptual knowledge of physics and their confidence in solving problems. Within the realm of problem solving, students further identified the coaches as helping them with getting

TABLE X. Results from a spring 2013 end-of-semester survey question asking students to rank (with no ties), ten ways in which the coaches were useful to them in order from most (1) to least useful (10). Lower numbers are better.

<i>L</i> group (<i>N</i> = 57) (light users)	<i>M</i> group (<i>N</i> = 37) (medium users)	<i>H</i> group (<i>N</i> = 44) (high users)
Applying the physics concepts to a specific problem 4.0 ± 0.4	Getting started solving a problem 3.8 ± 0.7	Applying the appropriate equations to a particular problem 4.0 ± 0.6
Deciding what physics to use (kinematics, conservation of energy, etc.) 4.0 ± 0.4	Applying the physics concepts to a specific problem 4.1 ± 0.5	Getting started solving a problem 4.1 ± 0.6
Interpreting the problem text 4.5 ± 0.6	Deciding what physics to use (kinematics, conservation of energy, etc.) 4.1 ± 0.7	Applying the physics concepts to a specific problem 4.4 ± 0.5
Applying the appropriate equations to a particular problem 4.5 ± 0.5	Applying the appropriate equations to a particular problem 4.3 ± 0.5	Deciding what physics to use (kinematics, conservation of energy, etc.) 4.5 ± 0.5
Getting started solving a problem 4.5 ± 0.6	Interpreting the problem text 5.1 ± 0.7	Interpreting the problem text 5.1 ± 0.6
Determining that you need outside help 6.1 ± 0.6	Gaining confidence solving problems 5.5 ± 0.5	Gaining confidence solving problems 5.9 ± 0.6
Gaining confidence solving problems 6.4 ± 0.5	Doing better on the quizzes 6.6 ± 0.5	Doing better on the quizzes 6.1 ± 0.6
Doing the math 6.6 ± 0.6	Understanding the lectures 6.9 ± 0.6	Determining that you need outside help 6.6 ± 0.6
Doing better on the quizzes 6.9 ± 0.4	Doing the math 7.2 ± 0.5	Doing the math 6.8 ± 0.7
Understanding the lectures 7.4 ± 0.4	Determining that you need outside help 7.4 ± 0.6	Understanding the lectures 7.5 ± 0.4

started solving a problem and identifying and applying physics concepts and principles to a problem.

An analysis of the characteristics of students who chose to use the coaches finds that they appear to be those less confident about their abilities at the beginning of the course. They expect to spend more time studying and to earn lower grades than students who chose to use fewer coaches. In addition, the FCI shows that the more frequent users of coaches begin the class with a worse intuition about forces and motion than infrequent users. It appears that students in this environment who feel potentially at risk are receptive and motivated to use a tool that seems valuable, nonthreatening, and easy to use. Finally, we observe that, relative to the class as a whole, female students tend to be overrepresented among the students who are frequent users of the coaches and underrepresented among students who are infrequent users of the coaches.

From the data presented here, we conclude that web-based computer coaches designed to emphasize the expert-like metacognitive aspects of problem solving using a modified Socratic approach to take students through the many decisions necessary to solve a problem can be constructed. Moreover, we have shown that such coaches will be used and valued by a significant subpopulation of introductory physics students. In a subsequent paper, we will present evidence that the use of such coaches results in significant gains in students' problem-solving performance, as well.

VI. IMPLICATIONS FOR FUTURE DEVELOPMENT

These computer coaches that emphasize the metacognitive decision-making aspect of problem solving are an initial step in designing a useful software framework to provide usable on-demand coaching to students. We believe that the primary function of Internet coaches should be to supplement human classroom teaching by providing students with enough guided repetition so that they become comfortable with practicing problem solving as a series of decisions. From the results of our study, it is clear that students recognize this need and that existing technology is capable of providing it to a large number of students. We have also shown that, when well integrated into a course, a significant subset of introductory physics students, those potentially at risk, will use such coaches without much incentive.

These prototype coaches are only a beginning because any viable software framework needs to satisfy multiple

stakeholders: students, instructors, and institutions. To move forward, the software framework must be flexible enough to provide coaching for a wide range of students in a manner that adapts to their intellectual growth during a course. In the prototype coaches, apparent flexibility for students was provided by predetermined branching within the code. Additional student flexibility was achieved by building coaches that interacted with students in three different ways, as described briefly in this paper. However, more flexibility is needed. On the midsemester survey in spring 2013, about half of the respondents ($57 \pm 4\%$) agreed or strongly agreed with the statement "The coaches were too repetitive." Using object-oriented programming, the next generation of coaches can provide students with more flexibility to choose their own solution paths, including ways to construct a solution more quickly once they gain some problem-solving competence.

Physics instructors can easily use the 35 existing coaches if no changes are desired. However, to be useful to most instructors, a coach must be easily adaptable to fit into their pedagogy and teaching style. When other instructors outside our research group have used the coaches, they have requested such changes. Modifying the prototype coaches requires the ability to program in the underlying software, Flex which, while not difficult, is a significant barrier. A graphical user interface that allows instructors to make significant changes in the coaches without software knowledge will allow the coaches to be used by a wide range of physics instructors. We note that the next generation of computer coaches allowing for more student and instructor flexibility is currently under construction.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the combined work of many faculty, graduate and undergraduate students, post-docs, and programmers. Special acknowledgments go to Fred Reif for inspiring the creation of these coaches with his pioneering work on the PALs at Carnegie Mellon University and to Pat Heller for her work on student problem solving that is reflected in these coaches. We also thank Jen Docktor, Paul Haines, Koblar Alan Jackson, Yuichi Kubota, Jia-Ling Lin, and Jie Yang for their important contributions, as well as the contributions of numerous undergraduates who coded the coaches and pulled and organized student usage data. This work was supported by the University of Minnesota and the National Science Foundation under Grants No. DUE-0715615 and No. DUE-1226197.

- [1] P. Nicholson, A history of E-learning, in *Computers and Education: E-Learning from Theory to Practice*, edited by B. F. Manjón (Springer Dordrecht, 2008), p. 1.
- [2] A. Molnar, Computers in education: A brief history, *THE Journal* **24**, 63 (1997).
- [3] R. Fuller, Resource letter CPE-1: Computers in physics education, *Am. J. Phys.* **54**, 782 (1986).
- [4] National Research Council, *Exploring the Intersection of Science Education and 21st Century Skills* (The National Academies Press, Washington, DC, 2010).
- [5] National Research Council. *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century* (The National Academies Press, Washington, DC, 2012).
- [6] J. Dewey, *How We Think* (D.C. Heath & Co., Boston, 1910).
- [7] National Research Council, *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (The National Academies Press, Washington, DC, 2012).
- [8] M. E. Martinez, What is problem solving?, *The Phi Delta Kappan* **79**, 605 (1998).
- [9] E. Yerushalmi, C. Henderson, K. Heller, P. Heller, and V. Kuo, Physics faculty beliefs and values about the teaching and learning of problem solving. I. Mapping the common core, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020109 (2007).
- [10] J. Smith, A. diSessa, and J. Roschelle, Misconceptions reconceived: A constructivist analysis of knowledge in transition, *J. Learn. Sci.* **3**, 115 (1993).
- [11] R. Millar and J. Osborne, *Beyond 2000: Science Education for the Future* (King's College, School of Education, London, 1998).
- [12] D. P. Maloney, Research on problem solving: Physics, In *Handbook of Research on Science Teaching and Learning*, edited by D. L. Gabel (Macmillan, New York, 1994).
- [13] L. Hsu, E. Brewster, T. Foster, and K. Harper, Resource letter RPS-1: Research in problem solving, *Am. J. Phys.* **72**, 1147 (2004).
- [14] J. Bransford, R. Sherwood, N. Vye, and J. Rieser, Teaching thinking and problem solving: Research foundations, *Am. Psychol.* **41**, 1078 (1986).
- [15] A. Newell and H. A. Simon, *Human Problem Solving* (Prentice-Hall, Englewood Cliffs, NJ, 1972).
- [16] *Education & Competitiveness: A Resource and Policy Guide* (Partnership for 21st Century Skills, Washington, DC, 2008); http://www.p21.org/storage/documents/21st_century_skills_education_and_competitiveness_guide.pdf.
- [17] J. L. Docktor and J. P. Mestre, Synthesis of discipline-based education research in physics, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020119 (2014).
- [18] J. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, Expert and novice performance in solving physics problems, *Science* **208**, 1335 (1980).
- [19] A. C. Graesser, S. Lu, G. T. Jackson, H. H. Mitchell, M. Ventura, A. Olney, and M. M. Louwerse, AutoTutor: A tutor with dialogue in natural language, *Behav. Res. Meth. Instrum. Comput.* **36**, 180 (2004).
- [20] K. VanLehn, P. W. Jordan, C. P. Rosé, D. Bhembe, M. Böttner, A. Gaydos, M. Makatchev, U. Pappuswamy, M. Ringenberg, A. Roque, S. Siler, and R. Srivastava, The architecture of Why2-Atlas: A coach for qualitative physics essay writing, In *Intelligent Tutoring Systems: 6th International Conference*, (Spring, Berlin, 2002), p. 158.
- [21] L. S. Myneni, N. H. Narayanan, S. Rebello, A. Rouinfar, and S. Puntambekar, An interactive and intelligent learning system for physics education, *IEEE Trans. Learn. Technol.* **6**, 228 (2013).
- [22] A. Papadimitrou, M. Grigoriadou, and G. Gyftodimos, Interactive problem solving support in the adaptive educational hypermedia system MATHEMA, *IEEE Trans. Learn. Technol.* **2**, 93 (2009).
- [23] F. Reif and L. A. Scott, Teaching scientific thinking skills: Students and computers coaching each other, *Am. J. Phys.* **67**, 819 (1999).
- [24] L. A. Scott, Ph.D. thesis, University of Pittsburgh, 2001 (unpublished).
- [25] K. VanLehn, C. Lynch, K. Schulze, J. A. Shapiro, R. Shelby, L. Taylor, D. Treacy, A. Weinstein, and M. Wintersgill, The Andes physics tutoring system: Lessons learned, *Int. J. Artif. Intell. Educ.* **15**, 147 (2005).
- [26] K. A. Ericsson and N. Charness, Expert performance: Its structure and acquisition, *Am. Psychol.* **49**, 725 (1994).
- [27] H. L. Dreyfus and S. E. Dreyfus, Expertise in real world contexts, *Organizational Studies* **26**, 779 (2005).
- [28] K. Heller, *The Competent Problem Solver for Introductory Physics* (Thomson, Independence, KY, 2006).
- [29] D. P. Maloney, in *Getting Started in PER, Reviews in PER Vol. 2*, edited by C. Henderson and K. A. Harper (Am. Assoc. Phys. Teachers, College Park, MD, 2011), <http://www.per-central.org/items/detail.cfm?ID=11457>.
- [30] B. Eylon and F. Reif, Effects of knowledge organization on task performance, *Cognit. Instr.* **1**, 5 (1984).
- [31] J. Larkin and F. Reif, Understanding and teaching problem solving in physics, *Eur. J. Sci. Educ.* **1**, 191 (1979).
- [32] M. Chi, R. Glaser, and E. Rees, Expertise in problem solving, in *Advances in The Psychology of Human Intelligence*, edited by R. J. Sternberg (Erlbaum, Hillsdale, NJ, 1982), Vol. 1, 7.
- [33] F. Reif and J. Heller, Knowledge structure and problem solving in physics, *Educ. Psychol.* **17**, 102 (1982).
- [34] E. Bagno and B. Eylon, From problem solving to knowledge structure: An example from the domain of electromagnetism, *Am. J. Phys.* **65**, 726 (1997).
- [35] R. Elio and P. Scharf, Modeling novice-to-expert shifts in problem-solving strategy and knowledge organization, *Cogn. Sci.* **14**, 579 (1990).
- [36] K. Cummings, J. Marx, R. Thornton, and D. Kuhl, Evaluating innovation in studio physics, *Am. J. Phys.* **67**, S38 (1999).
- [37] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. part 2: Designing problems and structuring groups, *Am. J. Phys.* **60**, 637 (1992).
- [38] A. Collins, J. Brown, and S. Newman, Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics, in *Knowing, Learning and Instruction*, edited by R. Glaser and L. B. Resnick (Erlbaum, Hillsdale, NJ, 1989), p. 453.

- [39] J. S. Brown, A. Collins, and P. Duguid, Situated cognition and the culture of learning, *Educ. Res.* **18**, 32 (1989).
- [40] D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).
- [41] A. L. Palincsar and A. L. Brown, Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities, *Cognit. Instr.* **1**, 117 (1984).
- [42] X. Zhu and H. A. Simon, Learning mathematics from examples and by doing, *Cognit. Instr.* **4**, 137 (1987).
- [43] R. A. Montgomery, L. Peguy, and M. Cannella, *The Abominable Snowman (Choose Your Own Adventure, No. 1)* (Chooseco Waitsville, VT, 2005), is one example of many such books.
- [44] E. Wenger, *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge* (Morgan Kaufmann, San Francisco, 1987).
- [45] G. Polya, *How to Solve It* (Princeton University Press, Princeton, NJ, 1945).
- [46] J. Heller and F. Reif, Prescribing effective human problem-solving processes: Problem description in physics, *Cognit. Instr.* **1**, 177 (1984).
- [47] P. Hardiman, R. Dufresne, and J. Mestre, The relationship between problem categorization and problem solving among experts and novices, *Memory Cognit.* **17**, 627 (1989).
- [48] L. Walsh, R. Howard, and B. Bowe, Phenomenographic study of students problem solving approaches in physics, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020108 (2007).
- [49] K. Schultz and J. Lochhead, *Toward a Unified Theory of Problem Solving: Views from the Content Domains* (Lawrence Erlbaum Associates, Hillsdale, NJ, 1991).
- [50] A. Van Heuvelen, Learning to think like a physicist: A review of research-based strategies, *Am. J. Phys.* **59**, 891 (1991).
- [51] J. Docktor and K. Heller, Assessment of student problem solving processes, in *Proceedings of the 2009 Physics Education Research Conference (AIP Conference Proceedings Vol. 1179)*, edited by M. Sabella, C. Henderson, and C. Singh (AIP, Melville, NY, 2009), p. 133.
- [52] P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping. part 1: Group vs. individual problem solving, *Am. J. Phys.* **60**, 627 (1992).
- [53] C. A. Ogilvie, Changes in students' problem-solving strategies in a course that includes context-rich, multifaceted problems, *Phys Rev. ST Phys. Educ. Res* **5**, 020102 (2009).
- [54] J. R. Anderson, A. T. Corbett, K. R. Koedinger, and R. Pelletier, Cognitive tutors: Lessons learned, *J. Learn. Sci.* **4**, 167 (1995).
- [55] R. E. Mayer, *Multimedia Learning*, 2nd ed. (Cambridge University Press, New York, 2009).
- [56] J. Sweller, Implications of cognitive load theory for multi-media learning, in *The Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer (Cambridge University Press, Cambridge, England, 2005), p. 1930.
- [57] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).
- [58] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, *Phys. Teach.* **30**, 141 (1992).
- [59] <http://groups.physics.umn.edu/physed>.
- [60] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. K. Dubson, D. C. Finkelstein, and C. Wieman, A new instrument for measuring student beliefs about physics and learning physics: The colorado learning attitudes about science survey, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010101 (2006).
- [61] P. A. Tipler and G. P. Mosca, *Physics for Scientists and Engineers* (W. H. Freeman, New York 2008).
- [62] M. Lorenzo, C. H. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, *Am. J. Phys.* **74**, 118 (2006).
- [63] S. J. Pollock, N. D. Finkelstein, and L. E. Kost, Reducing the gender gap in the physics classroom: How sufficient is interactive engagement?, *Phys. Rev. ST Phys. Educ. Res.* **3**, 010107 (2007).
- [64] J. Docktor and K. Heller, Gender Differences in Both Force Concept Inventory and Introductory Physics Performance, in *Proceedings of the 2008 Physics Education Research Conference (AIP Conference Proceedings Vol. 1064)* edited by C. Henderson, M. Sabella, and L. Hsu (AIP, Melville, NY, 2008), p. 15.