CoPractice: An Adaptive and Versatile Practice Tool

Kevin Buffardi,[1] Dzmitry Churbanau[2], Rahul Kanna N. Jayaraman[3], & Stephen H. Edwards [4]

Abstract – Intelligent tutoring systems and other adaptive instructional technologies facilitate learning by customizing feedback to students' struggles. However, adaptability of expert systems can come at the sake of versatility. Developing software to assist students in a particular domain can be challenging; to then apply its expertise and pedagogical approach to another domain can be both challenging and inappropriate. CoPractice is a web-centered practice environment that uses community-contributed content to establish a versatile system that adapts to learners based on measurements of their knowledge. Selection of both questions and feedback are based on persistent measurements of their effectiveness. This paper describes CoPractice's measurements and design for both adaptability and versatility.

Keywords: Practice, Assessment, Collaboration, Feedback Grading

INTRODUCTION

Practice is an integral device in promoting effectiveness in learning. Education researchers widely-acknowledge that "deliberate practice" elevates learning [1]; however, providing feedback to the learner is important for him or her to track progress, understanding, and application of the knowledge. Where static materials such as textbooks may only provide, at most, solutions to practice problems, leveraging technology adds the potential benefit of adaptive feedback.

Intelligent tutoring systems (ITS) are one type of pedagogically assistive technology that takes advantage of adaptive feedback. ITSs typically employ cognitive models to take form of expert systems. Beneficially, ITSs offer students an opportunity to "learn by doing" with individual interaction. Furthermore, ITSs do not need to have a human tutor available – making intelligent tutoring systems convenient for students to learn on their own when a human teacher or tutor is not available.

Koedinger and Sueker [8] introduced the Practical Algebra Tutor (PAT) which demonstrated success in assisting development of students' understanding of algebra as shown in improved standardized test scores [7]. PAT, and other second-generation tutors like it, uses artificial intelligence in practice environments to provide feedback (both positive and negative) and hints. A panel of ITS researchers concluded that such "second generation" tutors approach half the effectiveness of human tutors [3]. While the helpfulness without the intervention of human tutors is promising, the panel also acknowledged limitations and difficulties in trying to emulate human tutors with computer models.

Developing artificial intelligence to model and mold the learning process for students is not a simple exercise. Consequently, building ITSs faces the daunting challenge of implementing extensible expert systems. For example, it would not be easy to apply PAT's model of tutoring algebra word problems to tutoring non-mathematical subject areas. Therefore, there are benefits in designing a different kind of instructional technology with an objective to avoiding the rigidity of expert systems where it can apply across many subject areas.

---

[1] Virginia Tech, 114 McBryde Hall (0106), Blacksburg, Virginia, kbuffardi@vt.edu

[2] Virginia Tech, 114 McBryde Hall (0106), Blacksburg, Virginia, dzmitryc@vt.edu

[3] Virginia Tech, 114 McBryde Hall (0106), Blacksburg, Virginia, rahulknj@vt.edu

[4] Virginia Tech, 114 McBryde Hall (0106), Blacksburg, Virginia, edwards@cs.vt.edu

Accordingly, developers of intelligent tutor/courseware management system ELM-ART found that versatility was more important to teachers and administrators than any special feature [11]. This paper describes CoPractice, a practice tool designed particularly to support versatility without sacrificing adaptability. Instead of depending on artificial intelligence to establish a system's "expertise," CoPractice leverages the knowledge of a learning community to support topic-specific expertise.

RELATED WORK

CoPractice draws likenesses to other collaborative learning environments and practice systems. However, surveying the wide variety of automatic assessment tools available for supporting practice questions is beyond the scope of this paper. In comparing CoPractice to select peers, we make note to highlight its different approach in supporting adaptation and versatility.

PeerWise [4,5] is a multiple choice question online practice system that depends upon students to provide questions as well as feedback to the questions. In order to practice questions in the system, students are required to also contribute questions they created themselves. Although they are not required to, students may also provide feedback to their fellow students – written comments about the questions. PeerWise's authors cite literature reporting benefits in self-assessment and peer-assessment to suggest the system's support of engaging students in "cognitive demanding tasks" to reinforce their understanding.

While students receive feedback that other students offer for their questions, the drill-and-practice element of interaction does not adapt to the learner's ability or struggles. Instead, CoPractice solicits students to provide feedback (in the form of corrections or hints) to other students who practice that same problem. CoPractice's Feedback Selection Model (described in detail later) then chooses the best available feedback based on the question-taker's struggles.

While CoPractice does not offer written comments to the question's author, it does record students' subjective ratings. In addition, data collected on student-provided questions and feedback is available for the student to review. PeerWise also found a need for external motivators for students to generate content [6]. CoPractice offers additional motivation by recording students' "Teaching Score" (described later) based on how well their feedback assists others.

PeerWise's authors acknowledge limitations in multiple choice (as opposed to free-response) questions, but also describe the advantage of the relative low-cost of generating, assessing, and reusing multiple choice questions [4]. In fact, Traynor and Gibson [10] describe how multiple choice questions can be synthesized and analyzed autonomously, lowering the overhead of creating multiple choice questions further. While CoPractice does not exclude the possibility of adding automatically generated questions, it purposely supports student-generated questions in hopes to apply students' higher-order cognitive skills as described in a modern application of Bloom's Taxonomy of educational goals [2].

COPRACTICE

Overview

Before discussing the measurements and models implemented in CoPractice, we will illustrate the system's general organization, features, and interaction. Next, the details of the question selection, knowledge, feedback selection, and teaching models depict CoPractice's primary components. In these sections we discuss our design choices as well as the benefits and drawbacks of alternate directions in designing an adaptive and versatile practice environment.

CoPractice features three general tasks, each segmented into its own tab: Practice, Create, and Analyze. There are two user groups for the CoPractice system: students and teachers. By-in-large, the capabilities of the two roles intersect greatly with the exception of permissions in the Analyze tab, to be discussed later in this section. The majority of the features belong in the Practice tab, but first the system needs a question bank.

The Create tab is accessible by all users for the purpose of creating multiple choice questions for students to practice. Create involves form entry of the required information for each question: question text, correct and

incorrect answer options, category, difficulty, and default hint. The category serves as a tag to describe the topic of the question. The form allows the contributor to either select an existing category or create a new one.

We considered allowing more than one category to be assigned to a question. However, tracking the knowledge of a student for multiple different categories becomes more complex since the system would have to assess whether a particular wrong answer demonstrated a lack of knowledge for one or some combination of multiple categories assigned to that question. For that reason, CoPractice limits the author of a question to assign only a single primary category for the question.

The difficulty input in question creation is a five-point scale where the author estimates how hard the question is. CoPractice collects data from student practice and calculates its own estimation of the questions. The author can later review differences in question difficulty estimations in the Analyze tab.

Lastly, the default hint provides preliminary feedback for students who have trouble with the question. However, as more students practice and contribute to the system, their feedback will be added to supplement the default hint. We also have plans to support file uploading of questions so that the questions can be generated in bulk offline and then added all at once to the CoPractice system in future versions.

All questions are made available to the students via the Practice tab. When first entering the Practice tab, a student is presented with a table of links to each category available. The table also displays the student's Learning Scores and Teaching Scores for each category.

As students practice categories, leave, and return again, they can use the screen shown in Figure 1 to track how well they understand each category and how well their feedback is helping teach others per category. Tracking these scores may also serve in allowing instructors to monitor or use the scores for grading or incentives. Likewise, scores may indirectly introduce a social aspect to CoPractice, where students may gain extra motivation to practice in order to beat their peers' scores.

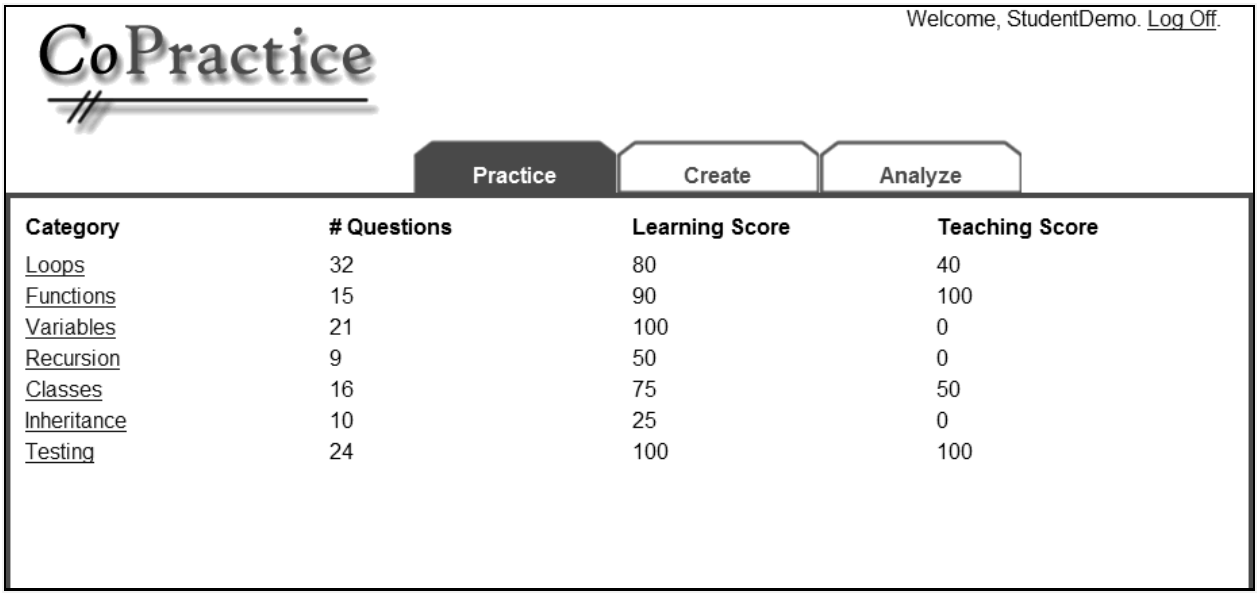

Figure 1. Screen capture of content within CoPractice's Practice tab with categories and affiliated scores.

Once a student selects the category he or she desires, CoPractice progresses to the drill-and-practice system. The student may continue to practice questions in that category as long as they want or until they: log out and quit, return to the list of all categories, or run out of available questions in that category.

In practice, the student chooses between the multiple choice answers in the right column of the screen after reading the question text in the left column. When the student's choice is incorrect, CoPractice shows a notification of the wrong answer, and provides the best available feedback (described in detail later). After reading the feedback, the student can attempt the question again. With each failed attempt at answering, the number of points available for the particular question decreases.

Students can also rate the feedback they received by choosing a "thumbs up" for good feedback and "thumbs down" for poor feedback. If the student chooses "thumbs down," he or she is provided with the next best feedback. However, students are discouraged from abusing the "thumbs down" feature and seeing all available feedback because with each feedback they receive, the fewer points they can receive for answering the question.

On the contrary, when a student answers a question correctly, CoPractice solicits the student to provide feedback for the question to help other students. At that time, the student's Teaching Score for the current category is shown and the student is encouraged to increase the score by providing helpful hints. However, feedback is voluntary and students may decide to skip to the next question for a more fluid drill-and-practice workflow.

Lastly, there is the Analyze tab. In the Analyze tab, students and teachers may review the data and calculations of feedback and questions. The one difference between the student and the teacher roles is the teacher may review all questions and feedback and delete at their discretion, while students may only do so for questions and feedback they created themselves.

In a table for analyzing questions, columns include: the question and category it belongs to; number of student attempts of the question; a breakdown of correct and incorrect responses; and a comparison of CoPractice's calculation of question difficulty, the author's estimated difficulty, and the average difficulty rated subjectively by students when attempting the question.

Likewise, the table for feedback analysis includes: the feedback and question to which it belongs; the number of times the feedback has been shown; the "thumbs-up" versus "thumbs-down" subjective rating; as well as the Helpfulness (as measured in the Feedback Selection Model, described later).

Question Selection Model

CoPractice selects questions to show based on its tracing of the student's knowledge and a combination of question difficulty and discrimination estimation and Item-Response Theory (IRT). The Graduate Record Examination (GRE) popularized the computer adaptive testing scheme of IRT and we built CoPractice's question selection based on Winter & Payne's extension of IRT [12].

Using this model, calculations are made based on outcomes of each question to estimate its difficulty and discrimination values. Discrimination describes how well the question differentiates students who have greater knowledge from those with lesser knowledge. A question with strong discrimination will evaluate all students as correct who have knowledge equal to or greater than the question's requirement. Likewise, those with knowledge less than what is necessary for the difficulty of the question will answer incorrectly. This example would result in the highest possible discrimination score of 1.

While Winters and Payne [13] record discrimination on a scale from 0 to 1, we use a scale of -1 to 1, where negative scores reflect questions of poor discrimination: students without the knowledge of the difficulty of the question are evaluated to be correct while students with the adequate knowledge are evaluated to be incorrect. Using this different scale for discrimination, we can suppress questions with negative discrimination because they evaluate student knowledge poorly – likely as a result of misleading wording or an incorrect answer key. However, we keep the data for poorly discriminating questions in the Analyze tab so that the author can then self-assess his or her question writing.

Difficulty measures the level of knowledge necessary to answer the question. When the question is first entered into the system, the author provides an initial difficulty value based on his or her subjective estimation. However, as more attempts to answer a question are made, CoPractice readjusts that question's difficulty. The composite difficulty level considers: the number of students attempting to answer, students' knowledge levels and question outcomes, the number of students who provide subjective rating of the difficulty (voluntary rating on a 1 to 5 star scale, more stars signifying higher difficulty), and the ratings they provided. The calculation does not inherently give greater weight to outcome-based difficulty but unless every student who attempts the question also provides a subjective star rating, a greater portion of data will represent question outcomes.

In accordance with these measurements, CoPractice records and continually refines calculations of each question's difficulty and discrimination factors with each attempt at answering it. Correspondingly, when a student fails at questions and their knowledge level drops, CoPractice selects the next question at a lower difficulty with the highest

discrimination possible. Similarly, as a student answers questions correctly, he or she receive progressively more difficult questions, with highly-discriminating questions getting precedence.

Knowledge Model

The model of tracing the student's knowledge is closely tied to the Question Selection Model. As suggested in the previous section, students' knowledge level continuously readjusts based on the outcome of each attempted question and the question's difficulty. CoPractice's Knowledge Model (in the form of a "Learning Score") bears some likeness to PAT's "knowledge tracing" [7] where a score reflects the student's progress in learning a topic from question to question. CoPractice benefits from the categorical organization of questions so that sequential practice questions all map to the same Learning Score.

Each question has a point value based on its difficulty level. If a student answers that question correctly, its point value is included in calculating the current category's Learning Score. If a student initially answers incorrectly the question point value decreases. With each incorrect attempt, the value decreases to reflect the odds of guessing the answer based on the number of remaining answers.

Correspondingly, the point value reflects an attempt to discourage abusing the feedback system. As the student requests more feedbacks, the point value for the current question decreases. While students may still try to see as many feedbacks as possible, they lose the possibility of gaining as many points as possible. Consequently, they are implicitly encouraged to consider each feedback they receive in an effort to answer the question with as little help as necessary.

Feedback Selection Model

A potential drawback of depending on mostly student-written feedback is that students may not be as pedagogically well-versed in offering help as an instructor or "expert." However, using student-provided feedback has multiple benefits. Firstly, providing feedback or hints to other students requires some self-assessment and meta-cognition in analyzing how one derived the correct solution. The Analyze tab also enables students to review each feedback and self-evaluate their helpfulness and how to better model future feedback. These activities engage students in higher-level cognition according to Bloom's Taxonomy [2].

Additionally, CoPractice uses post-feedback question outcomes to analyze feedback and consequently rank the feedback so better feedback for a particular question is given precedence. Ranking considers the outcomes of attempts after receiving said feedback ("Helpfulness") as well as the cumulative subjective "thumbs-up"/"thumbs-down" responses ("Rating"). Feedback that results in correct answers gets higher "Helpfulness" scores and is consequently more likely to be shown.

However, given the freedom in open-response feedback, there are possibilities for even good feedback to: adequately address all incorrect answers, address just one of the incorrect answers, or address some – but not all – of the incorrect answers. For this reason, CoPractice ranks feedback not only by question, but also by incorrect answer within the question. For example, in a question where the correct response is choice (D), one feedback may help the student comprehend why (A) and (B) are incorrect. A second feedback may help the student comprehend why (C) is incorrect. A third feedback may generally teach the concept required to understand why (D) is correct, while a fourth feedback is not effective in helping students recover from incorrect responses at all.

In this scenario, the first feedback (F_1 in Table 1) would have high scores for Helpfulness on (A) and (B), but low on (C). F_2 would have low scores for (A) and (B) but high on (C). F_3 would have high scores across the board, while F_4 has all low scores. In the case where a student incorrectly answers (C), he or she will first see either F_2 or F_3, determined by which Helpfulness score was higher for (C).

In situations where the Helpfulness scores are equal, the one with the higher Rating is chosen. Table 1 shows the scores for each of these feedbacks in the hypothetical scenario. Note that the (D) Helpfulness score for each feedback is 0 because no feedback will be provided after a correct response.

Feedback	(A)	(B)	(C)	(D)
F_1	+25	+30	0	0
F_2	0	-1	+15	0
F_3	+22	+19	+18	0
F_4	-1	0	-1	0

Table 1. Hypothetical Helpfulness scores for several feedbacks, in no particular order.

Teaching Model

Similarly to the relationship between Question Selection Model and Knowledge Model, the Feedback Selection Model and Teaching Model are closely tied. The Teaching Score is calculated using Helpfulness and Ratings of feedbacks. Like the Learning Score, the Teaching Scores are broken down by category so that students may see where their feedback for one category may be strong but for another category may be lacking. The Teaching Score for a single category reflects summation of Helpfulness scores for each answer for each feedback and as well as the corresponding subjective Ratings.

DISCUSSION

Online practice environments and automatic assessment are currently widely available. However, we believe CoPractice offers a novel approach that supports students' learning in: practicing, generating new questions and feedbacks, and in analyzing the results of the content they authored.

Additionally, CoPractice uniquely refines itself in improving both questions and feedback as the system is used. It automatically evaluates questions in terms of difficulty and discrimination and gives precedence to better questions and students are thereby less likely to come upon poorly discriminating questions or questions with difficulty far from their current capabilities.

Likewise, as the system collects more and more data on the feedback, the better CoPractice can predict feedback that best caters to the student's deficiencies in understanding the concept at hand. While the system already considers the favorability of the feedback and the effectiveness in the feedback helping students with their struggles, we plan to expand upon the feedback analysis. The best feedback ranking currently reflects feedback that students subjectively rate well and that helps the student derive at the correct answer. However, in future development, we plan to also evaluate how well the feedback helps the student learn the concept and apply it in other questions. In other words, a feedback that also helps a student correctly answer sequential questions in the same category will be preferred over a feedback that only helps answer the immediate question without demonstrating a deeper comprehension by also correctly answering related questions.

In addition, there is a need to improve CoPractice's implementation efficiency to support concurrent use by many users. Currently, most user actions result in immediate queries to the database to request or update information, including making calculations for each of the models. In order to support potentially large classes (or multiple sections of a class) with hundreds of students, we plan to ease the burden on the server, particularly in making some calculations scheduled on the server rather than recalculating upon every new data entry.

Finally, we plan to evaluate the current implementation of CoPractice in limited college classes where we can control across sections the treatment of having CoPractice available or not. In addition to comparing results in performance, satisfaction, and major retention across sections, we plan to also interview students and faculty about their impressions and recommendations after using CoPractice for a semester for formative evaluation.

REFERENCES

[1] J. D. Bransford, How People Learn, pp. 58-59, Washington, D.C.: National Academy Press, 2000.

[2] L. W. Anderson, D. R. Krathwohl, P. W. Airasian et al., A taxonomy for learning and teaching and assessing: A revision of Bloom's taxonomy of educational objectives: Addison Wesley Longman, 2001.

[3] A. Corbett, J. Anderson, A. Graesser et al., "Third generation computer tutors: learn from or ignore human tutors?," in CHI '99 ext. abstracts on Human factors in computing systems, Pittsburgh, Pennsylvania, 1999.

[4] P. Denny, J. Hamer, A. Luxton-Reilly et al., "PeerWise: students sharing their multiple choice questions," in Proc. of int. workshop on Computing education research, Sydney, Australia, 2008.

[5] P. Denny, A. Luxton-Reilly, and J. Hamer, "The PeerWise system of student contributed assessment questions," in Proc. of Australasian computing education - Volume 78, Wollongong, NSW, Australia, 2008.

[6] P. Denny, A. Luxton-Reilly, and J. Hamer, "Student use of the PeerWise system," in Proc. of Innovation and technology in computer science education, Madrid, Spain, 2008.

[7] K. R. Koedinger, J. R. Anderson, W. H. Hadley et al., "Intelligent Tutoring goes to school in the big city," International Journal of Artificial Intelligence in Education, vol. 8, pp. 30-43, 1997.

[8] K. R. Koedinger, and E. L. F. Sueker, "PAT goes to college: evaluating a cognitive tutor for developmental mathematics," in Proceedings of the 1996 int. conference on Learning sciences, Evanston, Illinois, 1996.

[9] J. J. McConnell, "Active and cooperative learning: final tips and tricks (part IV)," SIGCSE Bull., vol. 38, no. 4, pp. 25-28, 2006.

[10] D. Traynor, and J. P. Gibson, "Synthesis and analysis of automatic assessment methods in CS1: generating intelligent MCQs," in Proc. of the SIGCSE, St. Louis, Missouri, USA, 2005.

[11] G. Weber, and P. Brusilovsky, "ELM-ART: An adaptive versatile system for Web-based instruction," International Journal of Artificial Intelligence in Education, vol. 12, pp. 351-384, 2001.

[12] T. Winters, and T. Payne, "What do students know?: an outcomes-based assessment system," in Proc. of the first international workshop on Computing education research, Seattle, WA, USA, 2005.

[13] T. Winters, and T. Payne, "Closing the loop on test creation: a question assessment mechanism for instructors," in Proc. of SIGCSE, Houston, Texas, USA, 2006.

Kevin Buffardi

Kevin Buffardi is a Graduate Teaching Fellow and a Ph.D. candidate in Computer Science at Virginia Tech. He has an M.S. in Human-Computer Interaction from DePaul University and a B.S in Computer Science from University of Mary Washington. He also has multiple years experience in industry as a usability specialist. His primary research interests include: instructional technology, computer science education, and human-computer interaction.

Dzmitry Churbanau

Dzmitry Churbanau is a Masters student in Computer Science at Virginia Tech. He has B.S. in Computer Science and Applied Math from Belarussian State University. His research interests are in architecting and designing network-centric system of systems, theory of algorithms and novel programming paradigms.

Rahul Kanna N. Jayaraman

Rahul is a Masters student in Computer Science with research focus on Software as a Service and Network Simulations. He is also interested in building tools for online education and collaborative softwares. He worked in industry prior to joining for Masters and was instrumental in building web applications and collaborative softwares for Infosys Technologies.

Stephen H. Edwards

Stephen H. Edwards received the BS degree in electrical engineering from the California Institute of Technology, and the MS and PhD degrees in computer and information science from the Ohio State University. He is currently an associate professor in the Department of Computer Science at Virginia Tech. His research interests include software engineering, reuse, component-based development, automated testing, formal methods, and programming languages.