How Are Students' Problems Being Solved? The Quality of Worked Solutions on a Popular Open, Online, Mathematics Help Forum

Carla van de Sande, Arizona State University, carla.vandesande@asu.edu

Abstract: Free, open, online, help forums connect students with volunteer helpers who provide assistance with specific problems from coursework. One of the most popular existing mathematics help forums, Cramster, is an advocate of Cognitive Load Theory (CLT), and promotes the provision of worked solutions. The purpose of this project was to ascertain the quality of the worked solutions that students are receiving, by analyzing 128 threads from the Algebra thru Pre-Calculus site archives from two perspectives. First, a rubric based on research-based design principles of worked examples was used to distinguish poor from well-structured solutions. Second, student ratings of solutions they received were examined and compared with the research-based ratings. The results show that Cramster is providing students with responses that contain steps, and sometimes goals, as opposed to final answers only. However, there is also much misinformation in the responses, and students do not show much discernment in their ratings.

Introduction

Students often seek help understanding course material and completing their homework assignments outside of school hours. Free, open, online, homework help forums are an unregulated Internet-based resource that allow students to anonymously ask for help on specific problems using threaded discussions. Responses come from volunteers who have the time, desire, and (perhaps) experience to help. Many such forums exist for a wide variety of school subjects, and are especially popular for challenging, homework-intensive subjects like mathematics. One the largest forums, Cramster, can be found on www.cramster.com. Cramster is a proponent of Cognitive Load Theory (CLT), with influential educational researcher and theorist, John Sweller, as a member of its advisory board. Cramster's modus operandi is to give students access to worked solutions, and the site logo (CrA^{+}mster.com™, *problem solved*) reflects the conviction that this is a good – if not the best – way to help students learn. At the same time, Cramster recognizes that students are responsible for using the worked solutions they receive in a productive manner: "The Cramster.com team is comprised of current and former students and we know that while copying answers on Cramster.com will get your homework finished, it will not improve your understanding of a subject."

The purpose of this paper is to explore the nature of the worked solutions that students are receiving on Cramster in the context of Algebra thru Pre-Calculus. Due to the unregulated nature of the site, the responses might well consist of final answers only. On the other extreme, there is the possibility that, in accordance with Cramster's philosophy, the responses look very much like worked examples commonly found in textbooks or other instructional materials. As part of an ongoing investigation on help forum participation (van de Sande & Leinhardt, 2007, 2008; van de Sande, 2010, in press), this paper views the responses students receive on Cramster through the lens of cognitive research on worked-out examples, and also takes into account how the students themselves view the quality of the help they are receiving. To help situate the reader, a brief description of Cramster is first provided.

Cramster

Cramster is a global online study community for high school and college school subjects that was launched in 2003 by Cramster, Inc., a private company headquartered in Pasadena, California. Cramster offers spontaneous online help, in the sense that *any member can contribute* to the community, *regardless of their qualifications*. In addition to the Q&A board (homework help), the site provides access to study materials such as worked solutions to exercises found in popular textbooks, topic notes, sample problems, and practice exams that were created by the administration, indexed from the web, or contributed by members. Although general membership is free, participants enjoy additional benefits and access (e.g., ability to ask more questions per day on the Q&A board) if they subscribe monthly ($9.95/month) or annually ($49.95/year). The Q&A board has four sections (Math, Science, Engineering, and Business) with subforums organized by subject area (such as Algebra thru Pre-Calculus). Members have access to user profiles that include self-volunteered information on school, major, and expertise. In each contribution, members are characterized by a self-selected username, an optional avatar, board level (from Rookie to Oracle), and "Karma." Karma points are acquired through participation in site activities (such as answering queries on the Q&A board, submitting solutions to textbook problems, challenging others' solutions, providing notes, quizzes, outlines, etc., and referring friends to become members), and are based on an intricate point system. In the Q&A board, Karma points depend on the rating awarded by the student who asked the question (Lifesaver, Helpful, Somewhat Helpful and Not Helpful) and the difficulty level

of the board. For instance, answering a question on the Algebra thru Pre-Calculus forum posed by a non-paid/gold/platinum member could net 3/4/5 points, respectively, if the response receives a "Lifesaver" rating from the student who posed the request within 7 days. Similarly, Helpful and Somewhat Helpful ratings would net Karma points, just fewer. Karma points can be redeemed for gift certificates and merchandise, ranging in value from t-shirts to computers. On the part of students, rating responses is voluntary but does affect the student's "respect score" (the number of answers they have rated divided by the number of answers provided to them). This policy encourages students to rate the help received since students with low respect scores are presumably less likely to acquire (timely) help from forum helpers who are seeking to maximize their Karma points. Students can also leave comments, addressing a helper's response.

Cramster has recently (July, 2010) altered the look and feel of their site, together with many of their operative policies. The policies described above are the ones that were in effect at the time of the data collection. The current policies and their intentions were discussed in a phone interview with the Cramster press contact (Carleigh McKenna, personal communication, September 15, 2010). In the newer version of Cramster, instead of being able to post a specific number of questions per day depending on subscription type, students must earn or purchase Karma points in order to post questions. This policy discourages students from opening multiple accounts with different user names, and, at the same time, encourages students to actively contribute to the community (and thereby earn Karma points so that they can have their needs met). Students also now specify a time frame in which they need to receive a reply ("fast" for 12-24 hours, "faster" for 4-12 hours, and "fastest" for 0-4 hours), with each of these designated response speeds "costing" students progressively more Karma points. In addition to introducing variable point transactions for speed, the Karma point system was recalibrated, so that more Karma points (roughly 10 times as many) are at stake for a given service. Finally, the rating category "Somewhat Helpful" has been altered to "Needs More Work." The hope is that this change in semantics will encourage communication between students and helpers, so that helpers will make clarifications when needed and students can get additional help if the problem has not been resolved for them.

Theoretical Perspectives

Although help seeking once carried with it a stigma of laziness or incompetence, many theorists now recognize this activity as a potentially legitimate part of the learning process (Karabenick, 1998; Nelson-Le Gall, 1985; Newman, 1994), and one that underscores the Vygotskian (1978) notion of learning as inherently social. Whether or not help seeking has strategic value for learning depends on several attributes of the situation (Karabenick & Newman, 2006), including the performance goals of the participants. If the goals are mastery oriented, then the help-seeking encounter is instrumental and can provide a means of acquiring new knowledge and skills. On the other hand, if the goals are dependency oriented, then the help-seeking encounter is not a legitimate learning activity and instead merely serves as a means of acquiring information or solutions (Nelson-Le Gall, 1981).

It is within this theoretical perspective of help seeking that theories of learning come into play. Different learning theories offer us different definitions of what constitutes effective help and how best to promote mastery-oriented help seeking, particularly in the context of problem solving. Cognitive load theory (CLT), in particular, offers several extensively researched principles for the design of instructional materials (Mayer, 2005). Taking into account the limited capacity of working memory, CLT emphasizes the need for instructional design to minimize extraneous cognitive processing or load and to maximize germane load (Sweller, 1989; Sweller, van Merriënboer, & Paas, 1998). The argument is that certain practices contribute to extraneous cognitive load, making it more difficult than necessary for students to learn and retain the material, whereas other strategies lower extraneous cognitive load and allow all working memory capacity to be devoted to schema construction.

The use of worked examples is one practice that is consistent with reducing extraneous load, and that, therefore, CLT strongly advocates. The provision of worked examples as a study tool has been shown to facilitate near transfer, to shorten acquisition and performance time, and to make learning seem less effortful than engaging in problem solving or exploration (Paas & Van Merriënboer, 1993; Sweller & Cooper, 1985; Tuovinen & Sweller, 1999; van Merriënboer, Schuurman, De Croock, & Paas, 2002). These benefits are especially pronounced for students who are unfamiliar with the material that is being learned (Kalyuga, Chandler, Tuovinen, & Sweller, 2001).

However, worked examples are not all equally effective. The structure of worked examples matters, and "may substantially compromise the benefits derived from studying them" (Mwangi & Sweller, 1998, p. 174). Higher quality examples provide cues to students regarding the integration of various components (Atkinson, Derry, Renkl, & Wortham, 2000). In particular, examples in which goals and subgoals are highlighted, using labels or visual chunking, are more easily learned from than examples presented as a string of steps (Catrambone, 1995, 1998).

Methods

The Corpus

All of the threads from the archives of the Algebra thru Pre-Calculus Q&A board were collected for the month of September, 2009. This netted 1,246 threaded discussions. Next, in order to have a tractable number of threads to analyze regarding response quality, 10 threads from each day were randomly selected, resulting in 300 threads. Of these, 235 were deemed topic appropriate for Algebra thru Pre-Calculus. The Algebra thru Pre-Calculus threads were next coded for the type of response sought according to problem type: Solution (worked solution to given exercise), Explanation (explanation or description of concept), Clarification (clarification of solution retrieved from other source), Verification (assessment of accuracy for solution), and Information (formula). Because this project focuses on the quality of worked examples, only the 211 threads that involved a student seeking a *solution* were selected. There were 32 threads in which, against Cramster policy, a student posed more than one exercise, and these were discarded. Finally, only threads in which the problem statement, as presented by the student in the initial post, was well-defined (e.g., complete and unambiguous) were included. After this culling, the data set amounted to a total of 128 student requests for solutions and 194 helper responses.

The 128 student requests came from 82-94 different students, and the 194 responses came from 109-115 different helpers (since 13 students and 7 helpers used the shared user name "Anonymous"). All user names referred to in this paper have been altered to protect the (online) identity of participants, but an effort was made to use pseudonyms that still convey the user's persona.

The Analyses

All responses received to the 128 student requests for solutions were examined for quality from two perspectives: cognitive research-based principles for worked examples, and student perception.

Research-based Quality

To assess the quality of the help received on Cramster in light of theory-based design principles for worked examples, each helper response was assigned a rating from 0 to 5. Table 1 contains a description of some of the features that differentiated helper solutions. A 0 was assigned to helper responses that were either absent (0NR), were not a solution to the problem statement (0NS), or were mathematically inaccurate (0WR); a 5 was assigned to responses that almost had a "textbook quality" feel to them, clearly and concisely laying out the solution using steps, associated (sub)goals, labels, and, perhaps adding warnings or checks. Inter-rater agreement was 89% and all differences were resolved following discussion.

Table 1: Research-based quality rubric for solutions posted on Cramster.

Rating		Features
0	0NR	No response
	0NS	Response not representing solution to problem posed (e.g., solution to alternative problem statement)
	0WR	Mathematically inaccurate (wrong) response
1		Mathematically accurate response, but one that consists of a final answer only
2		Accurate response that provides a sketch of the solution, or an incomplete worked solution
3		Accurate response that closely resembles an instructor solution manual, containing some steps, but few, if any, (sub)goals or explanations
4		Accurate response that contains many steps and associated (sub)goals, but that is somewhat hard to follow (such as using inconsistent notation), or that contains extraneous steps (such as unnecessary conversions)
5		Accurate response that contains steps, associated (sub)goals, labels, and, perhaps, warnings or checks, with a clear and concise presentation style and format

Student-based Quality

Within 7 days of a response, students may rate the help on Cramster, as mentioned above. These ratings allow students to give the helpers feedback and provide an indication of how students view the quality of the responses. Because of the reformatting of the site, the original student ratings in the archives (whence the data set came) had been translated to the new rating system. There were 5 categories reflecting student pronouncement of quality, or lack thereof: Lifesaver (L), Helpful (H), Needs Work (NW) (old Somewhat Helpful), (NH) Not Helpful, and No Rating (NR).

Results

Research-based Quality

Figure 1 contains the percentage of responses according to their quality as worked examples using the research-based rubric. The quasi-bimodal distribution of the results tells the two sides of the story on the quality of the responses students are receiving on Cramster.

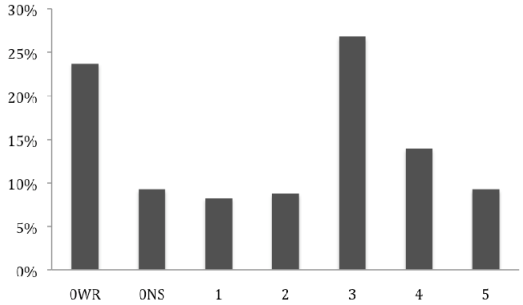


Figure 1. Percentage of Responses According to Quality as Worked Example.

The first part of the story is that students *are* receiving responses on Cramster, that the site is not just providing students with final answers, and, that many solutions posted are at least the quality of an instructor solution manual. Only 24 of the total 128 student requests in the data set, went unanswered (0NR). Although this is a relatively high nonresponse rate compared with similarly structured mathematics help forums (van de Sande & Leinhardt, 2007), it is still remarkable since Cramster helpers do not necessarily receive any compensation (Karma points) or commendation (forum status) for their efforts. The average number of responses to requests that were answered was approximately 1.9 (194 responses for 104 requests).

Furthermore, of the student queries that received responses, only approximately 7% of the responses consisted of a final answer only (1), resembling what one typically finds in a "back of the book" answer section. These responses are of absolutely no use as worked examples from which to learn, and violate Cramster's posted anti-cheating policy at that time: "Copying solutions or *posting unexplained final answers* on the Q&A Board promotes completion without comprehension, and that's something we don't support on this site" [italics added]. This low percentage is definite cause for comfort since the site is not regulated, standing outside the purview of formal educational institutions.

Finally, 44% of the solutions students are receiving on Cramster are the quality of those found in instructor solution manuals (3), or higher (4-5). These solutions, at the least, contain steps, and may also contain associated goals and explanations. Students could presumably use these solutions as worked examples to learn the material, much in the spirit of how they use examples that are provided in a textbook or other instruction materials. Figure 2 shows a high quality worked solution that was received in response to the problem statement: A 10-m ladder is leaning against a building. The bottom of the ladder is 5-m from the building. How high is the top of the ladder? The helper, Gondal, includes a labeled diagram, a goal ("use the pythaorean theorum"[sic]), an explanation ("right triangle"), and the solution broken down into steps. At the same time, this solution is not perfect; it contains several spelling errors, and, more importantly, does not explain why the negative solution value to the quadratic was not considered (e.g, because h is a length and therefore must be nonnegative). Despite these flaws, the student, What08, appreciated the quality of the response, giving it a Lifesaver rating and commenting "thanks."

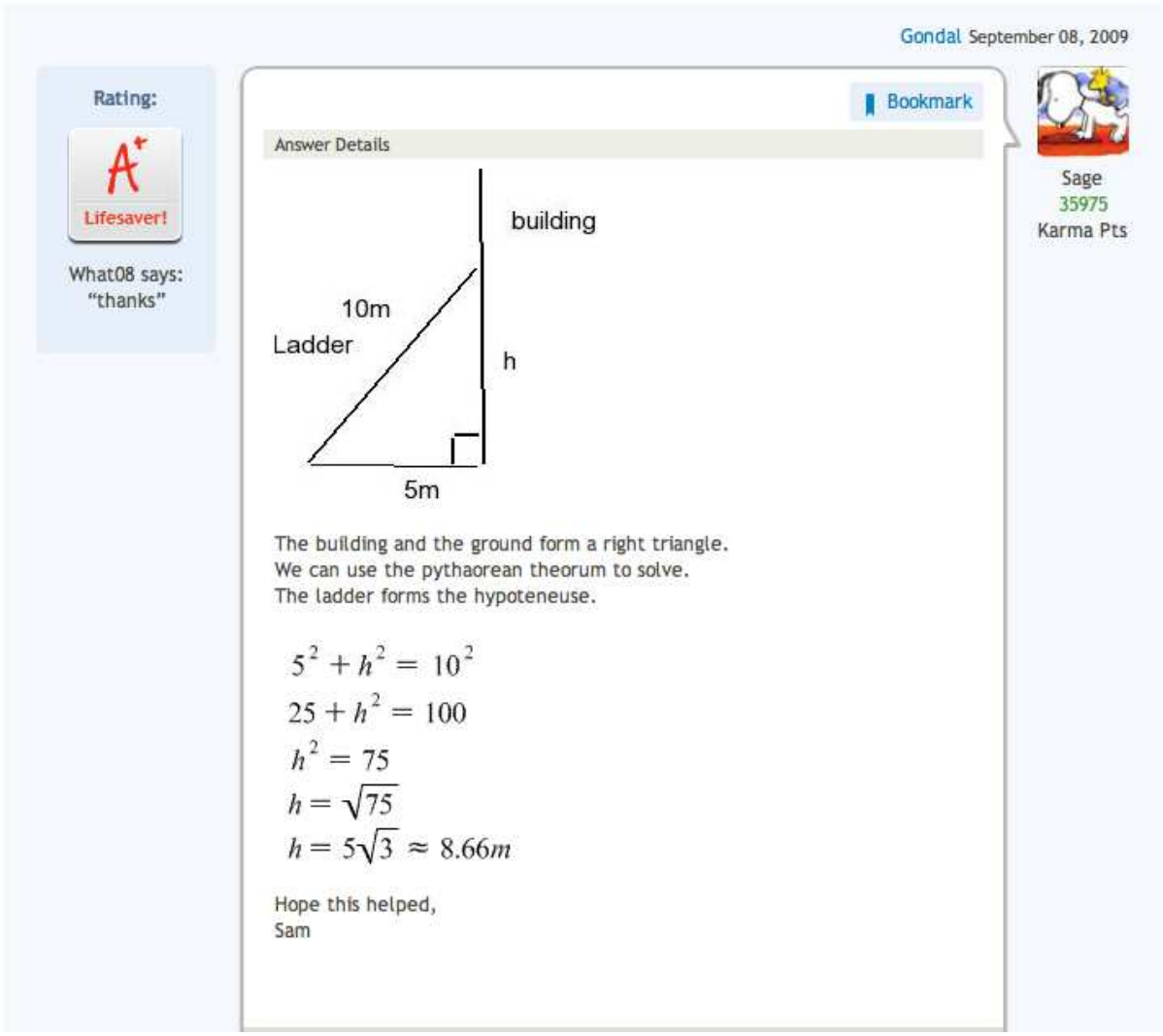


Figure 2. High Quality Worked Example Containing Steps, Explanations, and a Diagram.

The second part of the story, and a part that is cause for some concern, is that students are receiving incorrect information on Cramster, and that many of the responses they receive are not usable as worked examples. Indeed, 24% of the responses in the sample contained mathematically inaccurate statements (0WR). Here, I note that accuracy was coded very conservatively, and captured subtle errors that concerned mathematical aspects of the argument (such as the failure to use parentheses consistently), in addition to more flagrant transgressions (such as incorrectly factoring an expression). However, the justification for this decision was that the students who receive these responses cannot be assumed to have the mathematical experience necessary to detect errors (especially subtle ones) in an argument, and may be using these responses as a model.

Accuracy is the minimal criteria for being useful as a worked example. A number of responses also fell short of providing students with the semblance of a worked example. Altogether 41% of the responses received did not meet the criteria for being a passable worked example, either because they contained incorrect information (0WR), consisted only of the final answer (1), or, were in some way, incomplete (2).

Student-based Quality

Another metric for quality comes from the ratings given by student Cramster users to the responses they receive. Table 2 shows the number of responses (containing solutions) by student ratings (Not Helpful, Needs Work, Helpful, Lifesaver, No Rating) and the research-based quality rubric (0WR, 1-5).

As the last row in Table 2 shows, students did generally rate helper responses. Although they may have done so in order to maintain a high "respect score" (as discussed above), this score is completely independent of the actual student rating (a "Not Helpful" affects the student's respect score the same as a "Lifesaver" rating). Looking at Table 2, we see that, when they give them, students appear generous in their ratings. Of these responses that received ratings (n=163), the vast majority (n=156) received moderate (Helpful) to high (Lifesaver) marks, and no responses received the lowest rating, Not Helpful.

Table 2: Student ratings and research-based ratings for the quality of responses containing solutions.

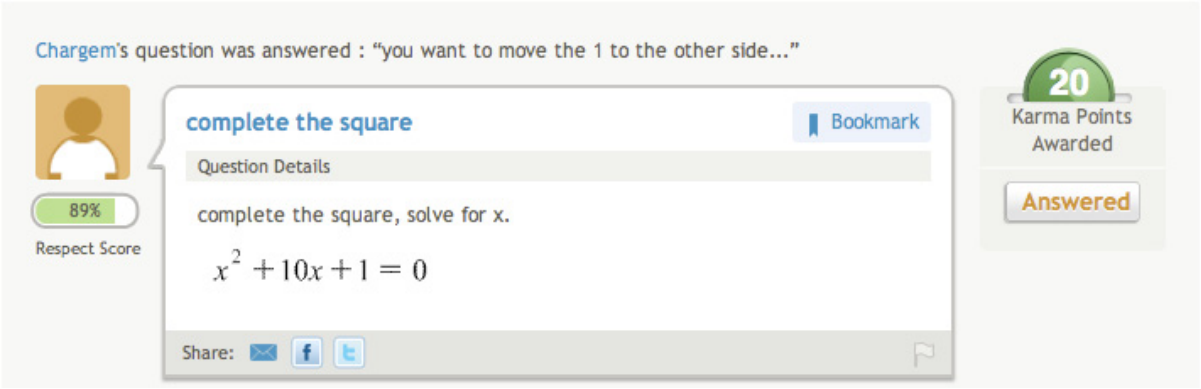
	Not Helpful	Needs Work	Helpful	Lifesaver	No Rating	Total
0WR	0	3	15	22	6	**46**
1	0	2	7	6	1	**16**
2	0	1	10	4	2	**17**
3	0	1	27	22	2	**52**
4	0	0	13	12	2	**27**
5	0	0	7	11	0	**18**
Total	**0**	**7**	**79**	**77**	**13**	**176**

Of course, the question that begs to be asked here is whether the marks are deserved, according to some independent and informed criteria, such as our quality rubric. The answer is that it appears that students are not very critical of "bad" help, and that they also appreciate "good" help. Of the solutions that were mathematically inaccurate (0WR), less than 7% (3 of 46) received poor marks. This trend also held for solutions that consisted of final answers only (1), for which only 13% (2 of 16) received a rating of "Needs Work." Solutions that were "good" worked examples (4-5) were recognized as such, receiving no poor marks (0 of 45) and equivalent Helpful (n=20) and Lifesaver ratings (n=23). The upshot is that students give high marks to inaccurate responses (55% received a Lifesaver rating, compared to 45% of the accurate responses). Also, students tend to be more prone to give higher ratings as the quality of the help increases.

An Example

Figure 3 shows an exchange in which an incorrect response received the highest rating possible and a correct response received a lower, although still good, rating. It is not clear what order the two responses were received in, since timing information, except for the date, is no longer published in the threads. The student, Chargem, gave the problem statement, which included a specific method, namely completing the square, and received a response from nzatisw58 that contained several quite serious mathematical errors. Indeed, the worked solution makes only one true statement (x^2+10x+1 = 0 can be rewritten as x^2+10x = -1). Chargem responded enthusiastically to this response, though, and, beyond giving it a "Lifesaver" rating, took the time to add an enthusiastic and appreciative comment for the "detailed explanation." Another response, from Firth, provided the correct solution (albeit with only one goal describing the first step for completing the square), but only received a "Helpful" rating without comment. Note, however, that neither helper alluded to a conceptual rationale for performing the steps that were included (e.g., a diagram illustrating the meaning of the terms).

Question:



Answers:

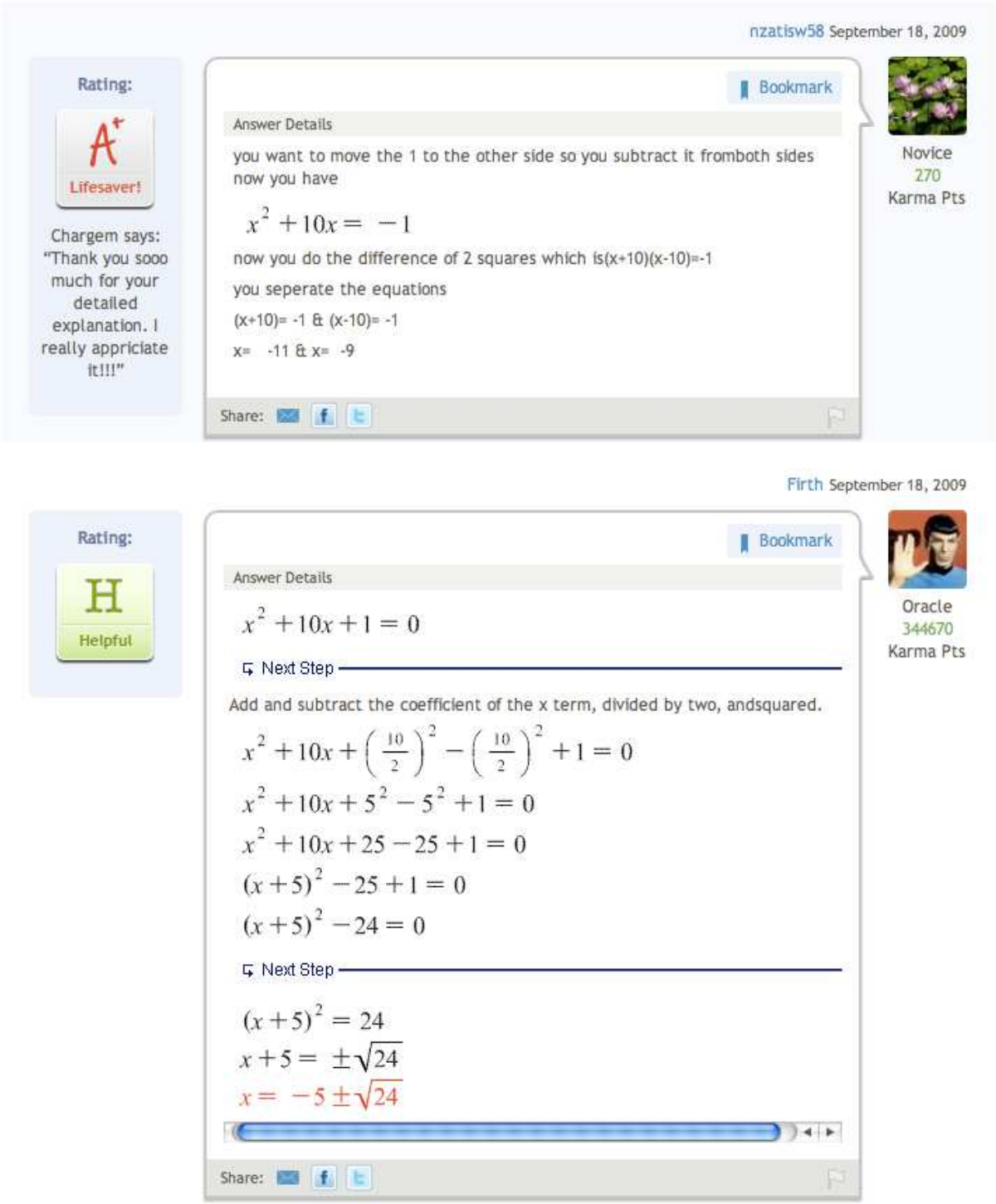


Figure 3. Question and Two Helper Responses: Incorrect one gets "Lifesaver." Correct response gets "Helpful."

Discussion

Thousands of students bring their problems to Cramster, presumably in the hopes of receiving a worked solution in good time from more experienced others. We see evidence from this project that students are generally receiving worked solutions as opposed to final answers, and that many of these worked solutions provide at least steps to follow. This is good news. On the other hand, many of the solutions also contain errors, some more serious than others. Such worked solutions could have the opposite effect than intended by giving unsuspecting and perhaps mathematically naïve students an incorrect example to follow as their guide. Based on their own ratings, it appears that students do not discern these errors and simply appreciate the fact that they have received assistance.

Of course, students may be communicating with helpers off-line (e.g., through private messaging) for clarification, and an observational methodology cannot capture this. Also, we cannot speak to the (ab)use of the

responses as worked examples, regardless of their quality. Are students using the solutions they receive as problem solving models, or are they passively copying them? In the former case, students would be experiencing the benefits of learning from worked-out examples, whereas, in the case of copying, students are cheating and no learning is taking place. Finally, we have no evidence that providing worked examples in an online help forum environment trumps alternative pedagogical practices (such as guided discovery) favored by other learning paradigms. In sum, this project addressed *how* student's problems are being solved – the question of whether students' problems *are* really being solved still awaits an answer.

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research, *Review of Educational Research, 70,* 181-214.

Catrambone, R. (1995). Aiding subgoal learning: Effects on transfer. *Journal of Educational Psychology, 87,* 5-17.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127,* 355-376.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Applied Educational Psychology, 93*, 579-588.

Karabenick, S. A. (1998). Help seeking as a strategic resource. In S. A. Karabenick (Ed.), *Strategic help seeking: Implications for learning and teaching* (pp. 1-11). Mahwah, NJ, Erlbaum.

Karabenick, S. A., & Newman, R. S. (Eds.). (2006). Help seeking in academic settings: Goals, groups, and contexts. Mahwah, NJ: Erlbaum.

Mayer, R. E. (Ed). (2005). *Cambridge Handbook of Multimedia Learning.* New York: Cambridge University Press.

Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction, 16,* 173-199.

Nelson-Le Gall, S. (1981). Help-seeking: An understudied problem-solving skill in children. *Developmental Review*, *1*, 224-246.

Nelson-Le Gall, S. (1985). Help-seeking behavior in learning. *Review of Research in Education*, *12*, 55-90.

Newman, R. S. (1994). Adaptive help seeking: A strategy of self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 283-301). Hillsdale, NJ: Erlbaum.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: an approach to combine mental effort and performance measures. *Human Factors, 35*, 737-743.

Sweller, J. (1989). Cognitive technology: some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology, 81*, 457-466.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59-89.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design, *Educational Psychology Review, 10(3)*, 251-296.

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*, 334-341.

van de Sande, C. (2010, June). *Free, open, online, mathematics help forums: The good, the bad, and the ugly*. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences 2010 (Vol. I, pp. 643-650). Chicago, IL: ISLS.

van de Sande, C. (in press). A Description and Characterization of Student Activity in an Open, Online, Mathematics Help Forum, *Educational Studies in Mathematics.*

van de Sande, C., & Leinhardt, G. (2007). Online tutoring in the Calculus: Beyond the limit of the limit. *Éducation et Didactique, 1(2),* 115-154.

van de Sande, C. & Leinhardt, G. (2008, June). *The Good Samaritan effect: A lens for understanding patterns of participation*. In P. A. Kirschner, F. Prins, V. Jonker & G. Kanselaar (Eds.), Proceedings of the 8th International Conference for the Learning Sciences 2008 (Vol. II, pp. 240-247). The Netherlands:ISLS.

van Merriënboer, J. J. G., Schuurman, J. G., De Croock, M. B. M., & Paas, F. G. W. C. (2002). Redirecting learners' attention during training: effects on cognitive load, transfer test performance and efficiency. *Learning and Instruction, 12*, 11-37.

Vygotsky, L. S., (1978). Mind in society: The development of higher psychological processes. Cambridge, Massachusetts, Harvard University Press.

Acknowledgments

This work was supported by a grant from the Spencer Foundation (#201000107).