Reinventing Homework as Cooperative, Formative Assessment

Don Blaheta
Department of Mathematics and Computer Science
Longwood University
Farmville, VA, USA
blahetadp@longwood.edu

ABSTRACT

Traditional processes for homework assignments are not always a good fit for the sorts of problems often seen in computer science classes. We present our experiences in implementing policies designed to encourage students to involve the instructor and fellow students in their learning process. By shifting to group assignments and permitting students a revision cycle, we improve student satisfaction and maintain or increase student outcomes while decreasing the instructor's grading load.

Categories and Subject Descriptors

K.3.2 [**Computers and Education**]: Computer and Information Science Education—*Computer science education*

Keywords

Formative assessment; Cooperative learning; Pedagogy

1. INTRODUCTION

Homework is a traditional, and often unexamined, tool in the educator's toolkit: guided practice to work on outside the classroom in order to develop skills and (hopefully) improve understanding. Though it may be graded work, we like to think of it as formative assessment, in the sense that it "helps students learn during the course of instruction."[5] The individual nature of homework, even among instructors who permit some degree of collaboration, is often seen as important precisely because of the opportunity it provides for individual practice and individual feedback.

The reality sometimes seems a little less rosy, however. We have observed that many students immediately file away graded work, spending little or no time reading instructor feedback; this greatly diminishes its value as formative assessment, of which feedback is a crucial component. D. Royce Sadler's earliest work on what he initially called "formative evaluation" was already enunciating an "iteration principle", that re-working an existing example seemed likely to be as important as, if not more important than, simply looking at an evaluation or trying to apply it to the next assignment.[4]

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SIGCSE'14, March 3–8, 2014, Atlanta, GA, USA.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2605-6/14/03 ...$15.00.
http://dx.doi.org/10.1145/2538862.2538915.

> Devise at least two interestingly-different tiebreakers for A* pathfinding on a 2D grid, and show a test case where they behave differently.
>
> Analyse which of your tiebreakers performs "best" on your test case, and discuss whether there is a tie-breaking strategy that will work well for all test cases or whether their relative performance depends on the problem.

Figure 1: A homework problem for an undergrad AI course (edited for length)

An additional problem with the more traditional homework model has to do with how it meshes with the idea of collaborative or cooperative learning among the students. To the extent that collaboration is permitted, any misconceptions held jointly by a group of students are not particularly well-addressed by separate, individual comments to each student. Individual students may believe they simply didn't quite understand what the group had talked about, or indeed that they might be the only one with the misunderstanding; this can hinder them from seeking out additional help or fixing the problem.

Although much of the work on formative assessment within computer science has focussed on developing new technology to support it (such as in [1, 2, 6]), we were interested to see how the existing tool—homework, using the classic "pencil and paper" technology—could be reimagined to better support the goal of formative assessment and cooperative learning. This idea seems similar in spirit to Tim S Roberts's work on adapting classically summative multiple-choice tests into a more formative assessment.[3]

In the spring of 2013, we taught an upper-level undergraduate elective on artificial intelligence. In that kind of course, typical homework questions (such as the one in Figure 1) often involve algorithm design and multiple points of subtlety, where the student will strongly benefit from interaction with the instructor and with fellow students. This paper reports on our experiences in that course, adapting our homework policy to provide a scaffold encouraging the use of student cooperation and instructor feedback to learn the course concepts.

2. REVISION CYCLES

As explained above, we had long observed that many students treat all assessment, including homework, as primarily summative rather than formative. Regardless of the type or quantity of comments written on their papers, many students would dutifully file away the graded homework, perhaps first looking at the overall grade, but spending little or no time on reading instructor feedback. Even many of the better students would read the feedback without really engaging with it. Furthermore, although students all had the ability to consult with the instructor during office hours or by email before the assignment was due, relatively few availed themselves of this opportunity. Students that got stuck early in a problem might never even begin to think about the more advanced material.

Our previous attempts to solve this problem involved simply letting students perform revisions after returning homework assignments. These efforts were largely unsuccessful. Many of the weaker students—who would most benefit from revisiting and revising their work—saw their original grade, deemed it acceptable, and ignored the revision opportunity entirely. Students who got good-but-not-perfect scores reacted similarly. The focus of many of the other students was on "recovering lost points" rather than really learning the material (which led to some frustrating teacher-student interactions). Moreover, the revision offer nearly doubled the amount of work for the instructor, since the original work was graded once, assigning partial credit as usual, and then the revision was graded again. The workload was not sustainable, especially in the face of negligible benefits.

The key insight that let us move forward on this problem was that we could use information asymmetry to shift the student's mindset away from points and grades: we can annotate the original submission with comments, but no scores. We developed this initial idea into a detailed plan surrounding revision-based homework:

1. Comments without grades
2. Revision cycle
3. Grades without comments
4. Limited scale

These four elements are each important to the plan, and work together in a few important ways.

Comments without grades. The professor hands back work with relatively detailed comments but no grades, while recording grades privately in case the original is lost or the student decides not to revise it. As noted, this is intended to take the focus away from their score, an extrinsic motivator and a source of stress, and to encourage a more internally-motivated, learning-oriented process.

Furthermore, from the perspective of the one writing the feedback, it is substantially easier to write helpful comments when not trying to simultaneously associate comments with individually gained/lost points from an analytic rubric. Rather than a defensive justification of lost points, the comments become constructive suggestions for improvement.

Revision cycle. Students have an opportunity to turn in a revised version of every homework problem. This element of the system is both central to the enterprise and also probably its least novel aspect. The "iteration principle" was part of Sadler's original description of formative assessment[4], and some form of revision or revisiting of assignments normally figures in any proposal for student-oriented formative assessment.

5: The answer is correct, or may have very minor errors in areas not addressed by the problem (e.g. simple arithmetic mistakes)

3: The answer demonstrates substantial understanding but is incomplete or contains errors in areas relevant to the problem.

0: The answer may or may not have included relevant facts, formulas, or figures, but demonstrates little or no clear understanding of how to apply them or approach the problem.

Figure 2: Rubric for homework problems

In the situation of homework with written comments, the particular advantage of the revision cycle is that it motivates all but the most detached of students to actually read and engage with the professor's feedback. It also addresses the classic weakness of one-shot graded homework, where two students who complete the course with full understanding may get different final grades: if one entered the course with some pre-existing understanding, under a traditional homework model that student "achieves a much higher points total. . . although their ultimate performances are equivalent."[4] If only the revision is graded, both students have the opportunity to achieve high final scores.

Grades without comments. Having received instructor comments on their original submission, on the revision the students will only receive a grade, without additional comments. This component has no direct advantage for the student, except inasmuch as it enables the rest of the system by not overloading the professor. The student has already received a full round of comments, though, so it is also not necessarily a disadvantage; and to some extent, this policy acknowledges the fact that many students would ignore them anyway, as many of them do in traditional single-round assignments. Students that wish to continue working on the material are, of course, encouraged to come to the professor's office hours.

Limited scale. The final component of the system is to simplify the rubric: scores would be assigned on a three-level scale as shown in Figure 2, rather than the more fine-grained per-problem analytic rubric scales we had used in previous courses. Rubrics with fewer distinct categories are a benefit to consistency and speed, with some cost to granularity. Such a pared-down rubric would probably be too harsh in an otherwise traditional system; but here, students whose *first* work demonstrated some understanding have a good opportunity to build on that and attain a 5 in the revision.

Furthermore, the limited scale contributes to the shift away from a grade-oriented mentality, as there is little opportunity to argue for small fractional points and no sense of "just missing" a particular score. It also contributes to the system as a sustainable teaching practice: assigning scores when only three levels are available is *much* easier. Together with the commenting/grading elements of the policy, which

ensure that the time-consuming full written feedback only occurs once, the more coarse-grained rubric ensures that the professor's workload is not nearly doubled. If anything, it may be somewhat lighter.

3. COOPERATIVE LEARNING

The other main thrust of our effort to improve students' engagement with homework assignments was to promote cooperative learning.

We had long been encouraging students to study together and discuss their assignments in order to improve their understanding and leverage their various strengths. However, their actual written work was expected to be done individually. This requirement had two somewhat negative effects. First, it discouraged the collaboration that we wished to promote, because students leery of cheating would prefer to avoid both the temptation and the appearance of working too closely with other students. Second, when students did collaborate, the requirement that they write up the work separately increased the grading load substantially, with limited benefit. Students working in a group would inevitably have similar problems with their answers, and so the grading would involve tediously writing out similar comments multiple times. While we believe that there was some benefit to requiring each student to compose their own answers, it's not entirely clear how large that benefit was, or whether it was worth the cost.

Thus, in conjunction with the policy changes discussed in the previous section, we decided to declare all homework assignments to officially be group work without restriction: students were permitted to work in groups of their own choosing, and all contributors would write their names at the top of the submission. In addition to solving or mitigating the problems described above, we hoped that this policy would work well in combination with the revision cycle to get students talking about the problems more than once and re-engaging with the concepts.

Such a policy is not without its potential problems, of course. Chief among them is the "free rider" problem: that without any rules or controls on the permissible level of collaboration, one or more students could receive credit for an assignment without having contributed to it substantially (or at all). In the extreme, one could imagine a single student doing the work for an entire group. A related worry was that the members of a group would split the work, each member separately doing one problem, but all receiving credit for all the problems. Before extending this policy to other courses, we very much wanted to see whether these problems would manifest in practice.

4. RESULTS

4.1 Early reactions

The early student reaction mostly focused on the group aspect and was fairly positive, although several students worried that there must be some sort of "catch". One asked if it would be within the rules for the entire section (about twenty students) to put their names on a single handin—which it was, although the logistical problems this would present for the students left us unconcerned that it would be a problem. In practice none of the groups that formed had more than five students.

Question	Response					Avg
	1	2	3	4	5	
Q1 Liked group work	0	0	2	3	8	4.46
Q2 Group work effective	0	0	5	2	6	4.08
Q3 Comment/revision effective	0	1	1	3	8	4.38

Table 1: Student responses to homework assignment format. (For text of survey questions and responses, see Appendix A.)

Anecdotal observation as the course progressed reassured us that our worst worries about the group work would not manifest. One group threatened to kick out a student unless that student began contributing, and in general it seemed that the students' attitude was to resent any freeloaders.

Another observation, also anecdotal, suggested that our willingness to *not* impose too much structure on the collaboration may have paid off in an unexpected way. One group developed a general workflow whereby they split up the problems and wrote them up separately for the initial handin; but then after receiving comments they met together to do the revision. They found it effective to have their "expert" on a problem explain it to the others as they read the instructor feedback and together worked out a revision. If we had imposed a structure on the cooperative aspect of the homework, this would not have been it; the students were able to make good use of their freedom to find a learning strategy that suited them.

4.2 Student response

At the end of the course, we asked the students to fill out a short survey about their experience. Out of 21 students who completed the course, $N = 13$ responded to the survey. Table 1 shows their responses to the homework policy: both changes met with overall student approval. The shift to group work was certainly popular (eight of the thirteen "really liked" this policy), but also seemed to them to improve their understanding: most respondents rated it as more effective at teaching the concepts in the course.

The effectiveness response was even stronger for the comment/revision policy: eleven of the thirteen responded that the revision opportunity made the homeworks more effective than a traditional homework format. In a free-response comment section of the survey, one student specifically observed that it "gave more motivation to actually read comments, and having the opportunity to address them definitely helped concepts sink in."

4.3 Outcomes

The size of the class ($N = 21$) was too small to draw any strong conclusions from the correlation of homework grades to exam grades, but we felt able to at least dispel the concerns about the "free rider" problem. Figure 3 on the next page shows a plot of the students' exam scores against their homework scores; the Pearson correlation coefficient of these scores is 0.66, a similar coefficient to those in this professor's other courses. More importantly, the distribution lacks any serious outliers, and the high-homework-low-exam quadrant is conspicuously empty. There appear not to have been any students who regularly attached their name to the group assignments without learning the material well enough to perform roughly as well on the exam as their group-mates.

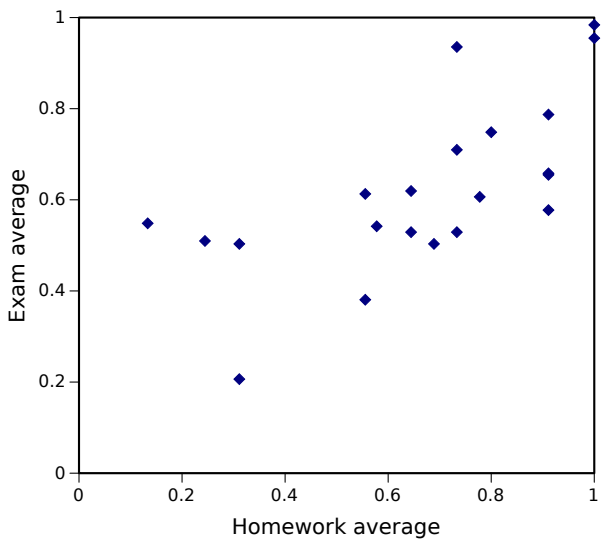

Figure 3: Comparison of homework and exam scores

The data shown here do not rule out the possibility that a student could join a group and passively absorb content without doing their fair share of work for the group, or that a student's pre-existing understanding coincidentally matches the products of the group they attach their name to. Neither possibility is a problem from a credentialing perspective, as long as the scores match, although they may be worth fixing for other reasons.

Another measure of the homeworks is particularly encouraging: Table 2 shows how the students responded to each revision opportunity. The table counts every problem that was handed in (once per group per problem), and tracks what that group subsequently did. Students were told that if they did not hand in a revision, their original grade (which was unknown to them) would stand. In just a few cases, a student or group decided not to revise their work, but in the vast majority of cases, at least minor revisions were made and handed in for the final grading.

About half of the 3s were improved to 5s, representing a substantial improvement. Most of the 0s were improved to 3s or even 5s as part of the revision. Students who did not hand in a problem are not counted here, so a zero in this table represents an actual attempt that was too incorrect to award even partial credit. Furthermore, this tactic engaged even the best students (those who were able to achieve a 5 on the first try); in general they performed minor revisions, but their second versions were more polished and often more thoughtful. At all levels, students made use of the revision cycle as an opportunity to pursue additional mastery of the topics.

5. CONCLUSIONS

Overall, the experience was definitely a success, and we are already using the new policies in other courses.

The students were nearly unanimous that this system is at least as good as a traditional homework model, with most finding it to be an improvement. The measurable outcomes do not contradict their assessment.

To the instructor (or grader) of a course, the policies decrease the overall grading load for a particular assignment. In addition, on a more subjective level, by decoupling the commenting from the grading, both aspects become somewhat easier. Knowing that the students are more likely to read the comments also helps.

Original score:	0	3	5
Count:	26	45	25

Original score	Revision score none	0	3	5
5	2			23
3	0		21	24
0	3	6	11	6

Final score:	0	3	5
Count:	9	32	55

Table 2: Frequency of scores on original submission and on revision

The system did introduce a minor logistical weakness: with only one copy of a homework for multiple people, only one person can be the custodian of the handed-back work. This was a minor problem for studying at exam time, especially as the work groups were not the same on every assignment. Photocopies were an adequate if somewhat ham-handed solution to the problem, although in the future a move to electronic submission and feedback should solve the difficulty more completely.

The kinds of high-level algorithm and data structure problems used in a course on artificial intelligence seemed to be an especially good fit for an assignment structure that encouraged working with other students and making use of instructor feedback. We look forward to testing their effectiveness in a variety of other computer science courses, and to collecting more student response data in order to draw stronger conclusions about the policies.

6. REFERENCES

[1] C. A. Higgins and B. Bligh. Formative computer based assessment in diagram based domains. In *Proceedings of the 11th Annual Conference on Innovation and Technology in Computer Science Education*, Bologna, 2006.

[2] A. Kyrilov and D. C. Noelle. Automatic formative assessment of exercises on knowledge representation in first-order logic. In *Proceedings of the 18th Annual Conference on Innovation and Technology in Computer Science Education*, Canterbury, 2013.

[3] T. S. Roberts. The use of multiple choice tests for formative and summative assessment. In *Proceedings of the Eighth Australasian Computing Education Conference (ACE2006)*, Hobart, 2006.

[4] D. R. Sadler. Evaluation and the improvement of academic learning. *The Journal of Higher Education*, 54(1):60–79, Jan–Feb 1983.

[5] L. A. Shepard. Formative assessment: caveat emptor. In *The future of assessment: shaping teaching and learning, ETS invitational conference 2005*, New York, 2005.

[6] A. Solomon. Linuxgym—software to automate

formative assessment of unix command-line and scripting skills. In *Proceedings of the 12th Annual Conference on Innovation and Technology in Computer Science Education*, Dundee, 2007.

APPENDIX

A. TEXT OF SURVEY QUESTIONS

Q1. Did you like or dislike having the ability to work with (and receive the same grade as) a group of people on the homeworks?

Choices for Q1: Really disliked / Somewhat disliked / Neither liked nor disliked / Somewhat liked / Really liked

Q2. Compared to a traditional do-it-on-your-own format, did you find that the group format for homeworks made them more effective or less effective at teaching you the concepts?

Q3. This term I introduced a homework format where I first wrote comments (but didn't tell you the grade), then permitted a revision, then assigned a grade (but didn't provide further comments). Compared to homeworks where comments and grade are both given at the same time (and no revisions allowed), was this course's homework format more effective or less effective at teaching you the concepts?

Choices for Q2 and Q3: Much less effective / Somewhat less effective / About the same / Somewhat more effective / Much more effective