

Chapter 1

Learnersourcing Debugging and Design Hints

This chapter presents two workflows that prompt students to engage with the diversity of solutions written by fellow students, reflect on their own debugging successes, and write hints that may help current and future students. It is adapted from a paper presented at the ACM conference on Computer-Supported Cooperative Work and Social Computing (CSCW) in 2016 [?]. References to “the authors” in this chapter refer to the coauthors of this paper.

One-on-one human tutoring is a costly gold standard in education. As established in Bloom’s seminal work on tutoring, mastery-based instruction with corrective feedback can offer a substantial improvement in learning outcomes over conventional classroom teaching [?]. However, personalized support does not scale well with the number of students enrolled. In large classes, it is often not feasible for students to get personalized hints from a teacher in a timely manner. Massive open online courses (MOOCs) can have even worse teacher-to-student ratios, by orders of magnitude. Intelligent tutoring systems have strived to simulate the type of personalized support received in one-on-one tutoring, but they are expensive and time-consuming to build.

~~In this chapter, students are asked to generate personalized hints for each other.~~ The two workflows in this chapter address this problem with a form of *learnersourcing*. Prior work on ~~*learnersourcing* demonstrates how~~ *learnersourcing* demonstrates that students can collectively generate educational content for future students, such as video outlines and exam questions, while engaging in a meaningful learning experience themselves [? ? ?]. ~~The proposed benefit of learnersourcing is that learners are not only more intrinsically motivated to engage with the learning content to begin with, but may also benefit pedagogically from the task itself.~~

This work builds upon learnersourcing by exploring how it can be applied to the generation of personalized hints during more complex problem solving. While prior work determined which learnersourcing task to present to the student based on what information the *system* needed [?], many educational topics like digital circuit design require some domain expertise that students are not equally well versed in. This raises the question of learnersourcing task routing. More specifically, which learners should be assigned to generate which hints? This work takes on the challenge of generating content that is tailored to both the *hint-receiver's* current progress and the *hint-author's* likely level of understanding.

This chapter ~~describes~~ presents two workflows for learnersourcing hints that assign hint-generating tasks to students based on what problems the student has recently solved. In the *self-reflection* workflow, students generate hints by reflecting on an obstacle they themselves have recently overcome. In the *comparison* workflow, students compare their own solutions to those of other students, generating a hint as a byproduct of explaining how one might get from one solution to the other. The second workflow can operate on the output of the first, as shown in Figure ?? . In both workflows, the key insight is that, through their own experience struggling with a particular problem, students can become experts on the particular optimizations they implement and bugs they resolve. The workflows can take pressure off teaching staff while giving students the valuable educational experiences of reflection and generating explanations.

While such workflows could have many applications, this chapter ~~presents a specific~~

~~application within a~~ describes their deployment in a college-level computer architecture classcourse. In this course, students implement, debug, and optimize simulated processors by constructing digital circuits. During both the debugging and optimization process, hints are one mechanism for teachers to help students fix and optimize their circuits. ~~This paper applies our learnersourcing workflows~~

The workflows are applied to two kinds of hints: *debugging hints* and *optimization hints*. A debugging hint is a student's attempt to help a future student change their solution so it generates the expected output for a particular input. An optimization hint is a student's attempt to help a future student get from one correct solution to another, more optimal solution. When hint-receivers encounter that particular situation during their problem-solving process, the hints can be shown to them as if they are the personalized hints an intelligent tutoring system might generate, or that a teacher might provide during a one-on-one interaction. The system for learnersourcing debugging hints is called Dear Beta and the system for learnersourcing optimization hints is called Dear Gamma.

1.1 System Design

There are a number of necessary decisions to make when designing interventions to collect and deliver hints. This chapter provides some answers to key design questions in the context of two systems deployed within a computer architecture class: ~~Dear Beta for learnersourcing debugging hints and Dear Gamma for learnersourcing optimization hints.~~

1.1.1 Design Questions

When should students be asked to provide hints? As soon as a student has resolved a bug, they may have some expertise about that bug that they can share with other students. If too much time has passed between resolving the bug and writing a hint, the student may forget necessary details and context, or forget the bug altogether. A previous learnersourcing

system also prompted students to contribute content immediately after having experienced it themselves, for similar reasons [?].

How should students access hints? Hints can be distributed using either a *push* or *pull* model, and can involve displaying either *all* or *some* of the hints. For example, a push model might display hints as a constantly updating resource, whereas a pull model could dispense hints to individual students just-in-time, when a student requests help. The hints could be algorithmically selected based on the student's work so far and the hints they have already received. This chapter demonstrates both the push and pull models, using the push-all model for distributing debugging hints and the pull-some model for distributing optimization hints. Generating optimization hints was a required reflection activity, so the volume and redundancy of these hints made a push-all model potentially overwhelming.

What hints can the system designer ask, or allow, a student to generate? In cases where the student *start state* prior to overcoming an obstacle and *end state* after overcoming the obstacle are known or inferred, the system can ask the student to write a hint that would help someone else overcome the same obstacle. For example, in the case of circuit design, a student who has recently fixed a bug by resolving a particular *autograder error* may be capable of writing a debugging hint for other students who get that autograder error. An autograder error occurs when a student solution does not return the expected values.

In many cases, however, a student might not face any explicit obstacle, or their start state may not be known. For example, a student might naturally arrive at a highly optimal circuit design without having first tried a less optimal design. Regardless of the path to their solution, the student could generate hints by comparing their own solution to a more optimal solution, or to a less optimal solution. This chapter explores both of these directions by asking each student to do both comparisons. To keep hint generation relevant to the student's current task and to minimize cognitive load, students were not asked to generate hints between pairs of solutions when they were familiar with neither solution.

How should hints be indexed? There are many possible ways to index the hints for more accurate retrieval later by students in need. For example, debugging hints could be indexed

by the autograder error they are intended to resolve. Optimization hints could be indexed by the student start state, end state, or both with respect to performance. Indexing hints by a meaningful feature of student solutions allows students to more easily find relevant hints in a push model of hint distribution and allows the system to deliver more relevant hints to each student in a pull model of hint distribution.

~~During debugging, student solutions are run through sequences of teacher-designed~~ The authors made specific indexing choices in order to build Dear Beta and Dear Gamma for students in the computer architecture course. In that course, students run their buggy solutions through teacher-designed sequences of test inputs. ~~Because test cases have a specification of actual and expected outputs for each input, we decided to index~~ Dear Beta indexes debugging hints by the ~~tests for which the solution deviates from the expected output. In other words, debugging hints are associated with the~~ first autograder error that disappears ~~if~~ once the bug is resolved.

During optimization, the goal is not simply to attain a correct solution, but ~~rather~~ to arrive at a more optimal correct solution. ~~We decided to index~~ Dear Gamma indexes optimization hints by both start and end states: ~~-,~~ which is the leap from a less optimal solution to a more optimal solution that the hint is intended to inspire. These states, the solutions themselves, are complex circuit objects. ~~We use the~~ The number of transistors in a solution as a metric of its optimality. In this indexing scheme, all hints written with the intent of helping a student with a 114-transistor solution create a 96-transistor solution are binned together.

Which hints should a student receive? In the push-all model of hint distribution, this question is not relevant, but in the pull-some model of hint distribution, it may be critical. Ideally, students would receive a progression of increasingly specific hints, following patterns of adaptive scaffolding established in intelligent tutoring system literature [?], helping them reach a more correct or optimal solution on the spectrum.

This ideal still leaves room for design choices about exactly which hints to deliver in a pull model of distribution. For example, during the optimization stage of an assignment, the system could *always* give the student hints that help them create the *next optimal solution*

found by other students. This would hopefully be an optimization challenge that falls within the student's zone of proximal development [?]. However, this may be too cautious. If a student's solution is far less optimal than the most common solution, then the system could give the student hints to help them leap directly to that most common solution, without first creating any intermediate solutions. This strategy ignores how large a leap outside their zone of proximal development this might be, but it ensures that the student is exposed to the ideas, presented in those hints, that are necessary for implementing a solution at least as optimal as the most common one. ~~We chose~~ Dear Gamma uses this latter option, both hinting students toward the most common solution and hinting students with the most common solution toward the most optimal solution.

Should hint creation be a required task? As discussed in Related Work, generating hints can be a valuable part of the learning process. ~~We required all students~~ All students were required to generate optimization hints as part of a reflection activity immediately after submitting their first correct circuit. ~~We did not require students to generate debugging hints. This is because the number of bugs encountered could be large, and unlike optimizations, many bugs also may not lead to significant conceptual gain upon reflection. Because debugging hints are immediately pushed out to all students, we keep hint creation voluntary to minimize the signal-to-noise ratio in hint quality. In contrast, writing debugging hints was voluntary. Some bugs, like random typos, do not warrant required reflection.~~

How should the variation in hint quality be handled? ~~Allowing students to upvote hints they find helpful provides one signal of quality. When available, we used student upvotes to sort hints. When we required every student to independently generate hints and there was no affordance for upvoting helpful hints, we presented five~~ When it is possible to collect data on helpfulness, collect it. For example, in a user interface, students can upvote hints and hints can be sorted by student upvotes. When hints are being pulled by students in need and there is little or no data about hint quality, the system could deliver multiple hints at a time to students in need of help. Even if one or more hints ~~were in the delivered batch are~~ were in the batch are of poor quality, their aggregate message might still be helpful for a student. If most of the hints ~~were in the batch are~~ about a feature that is irrelevant to the student receiving the hints, the

remaining hint(s) might be about something different and more relevant. ~~We limited each set of hints to a size of five to avoid overwhelming the student with too many hints~~Limiting the size of the batch to, for example, five, may prevent students from feeling overwhelmed.

How public should the hint-author be? Many systems for question-answering have a reputation system, where the author is known and recognized for contributing answers. Previous work at CSCW has examined whether reputation would improve class forum participation [?]. For simplicity, ~~we chose to leave student identities hidden~~Dear Beta and Dear Gamma both hide student identities.

1.1.2 Self-Reflection Workflow

In the self-reflection workflow, students iteratively modify their solutions to pass as many teacher-created tests as possible. For any autograder error revealed by those tests, students can look up hints for what modifications might cause their solution to pass that test case instead. The hints are stored in a database indexed by the autograder errors they are intended to address. When students fix a bug in their own solution, they can reflect on their fix and contribute a hint to the database for others struggling with the same autograder error. The self-reflection step turns a successful bug fix into a shareable hint. It is effectively a self-explanation exercise, since the hint is not a conversation starter. It is meant to stand alone for any future student to consult. As studied in prior work [?], self-explanation helps students integrate new information into existing knowledge.

1.1.3 Comparison Workflow

In the comparison workflow, students compare their correct solution to alternative correct solutions previously submitted by other students or teachers. They are prompted to compare their solution to a solution W with worse performance and generate an optimization hint for students who have solutions like W . They are then prompted to compare their solution to a solution B with better performance and generate an optimization hint for students who have

solutions like their own, which are not yet as performant as solutions like *B*. In each case, the student is generating an optimization hint to help students increase the performance of their solutions. Students who receive these hints use them as guidance while optimizing their own solutions. Figure ?? illustrates both workflows.

1.2 User Interfaces

~~We designed two user interfaces, one for each workflow.~~ To learnersource debugging hints with the self-reflection workflow, ~~we~~ the authors built and deployed *Dear Beta*, a Meteor web application that serves as a central repository of debugging advice for and by students in the class. The name alludes to both the “Dear Abby” advice column and the Beta processor that students create in the class. To learnersource optimization hints with the comparison workflow, ~~we~~ the authors designed *Dear Gamma*, a web interface students were required to fill out as a reflection activity after submitting their final correct circuit for a class assignment. Dear Beta and Dear Gamma each have their own user interface.

1.2.1 Dear Beta

~~We applied~~ Dear Beta is an application of the self-reflection workflow to processor debugging. Consider students working on their Beta processors within the digital circuit development environment provided by the computer architecture class. Students run a staff-designed test file x on their circuit. The development environment alerts them to a autograder error: for a particular input (test number n), y was the expected output and the student’s circuit returned z .

Students eliminate autograder errors by fixing bugs in their circuits. They may use trial and error, methodically examine internal simulated voltages, or experience a flash of insight. On the Dear Beta website, they can post a hint for others derived from their insight and indexed by the autograder error it caused. In the process of creating a hint, students have a

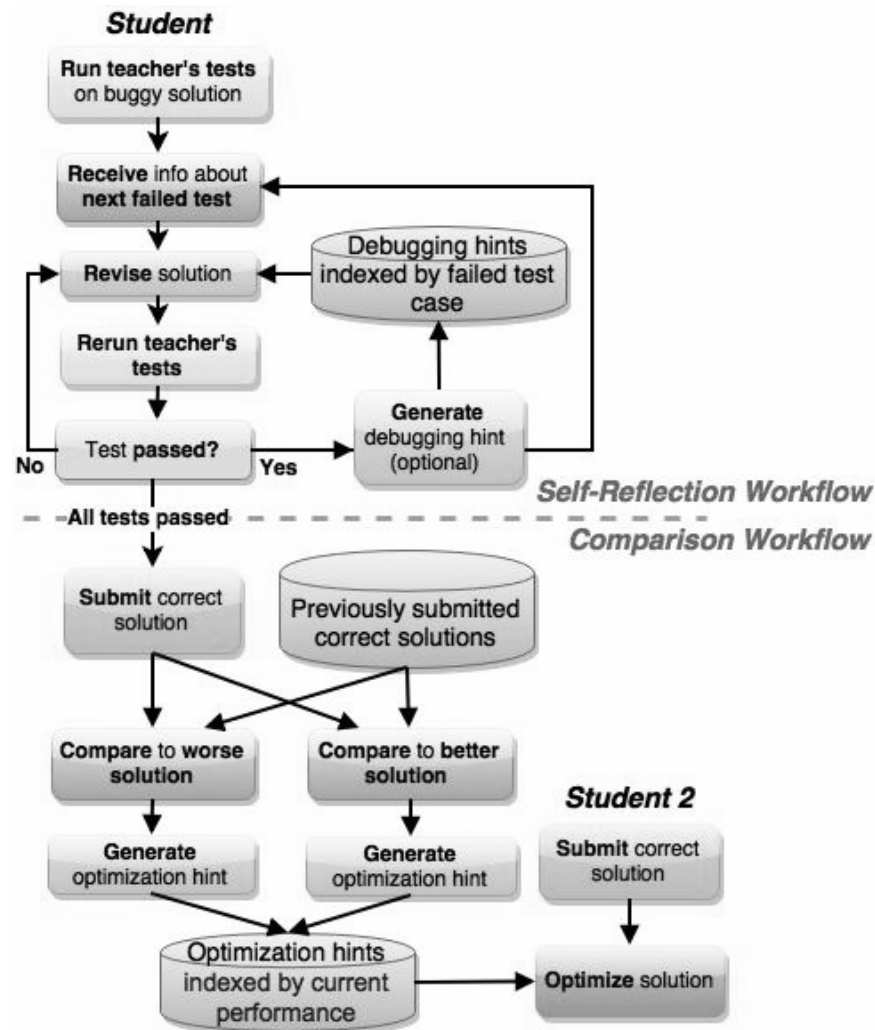


Figure 1-1: In the *self-reflection* workflow, students generate hints by reflecting on an obstacle they themselves have recently overcome. In the *comparison* workflow, students compare their own solutions to those of other students, generating a hint as a byproduct of explaining how one might get from one solution to the other.

chance to reflect on their own process of resolving the error. Other students encountering that autograder error can look up these hints, upvote helpful hints, and contribute new hints. Hints for each error are sorted by the number of upvotes they receive.

After further examining their malfunctioning processor with no success, some students may open the Dear Beta website in order to get help (Figure ??). Dear Beta displays all errors and hints, sorted by test file name x , then test number n . The student can either scroll to find hints for their autograder error or jump directly to them by entering x and n into the bar pinned to the top of the page. If the error was not yet in the Dear Beta system, the error will be added and scrolled into view.

If there are no hints yet, or if the hints are unhelpful, the student can click the “Request” button to the left of the error, which increments a counter. This button helps communicate the need for hints for a particular autograder error to potential hint writers. This is analogous to the “Want Answers” button and counter on Quora, a popular question-and-answer site.

If the student resolves their autograder error and feels that the existing hints were insufficient or incomplete, they can click in the text box labeled “Add a hint!” so that it expands into a larger textbox sufficient for typing out a paragraph of their own hint text (Figure ??). Their new hint may be a clearer rephrase of an existing hint or hint at yet another way to resolve the autograder error. Given the variety of processor designs and implementations, there may be several ways any given autograder error may be thrown.

Teacher Feedback on Early Prototypes of Dear Beta

After deploying initial prototypes of Dear Beta for two semesters, ~~we invited Teaching Assistants~~ Teaching Assistants were invited to share their complaints, requests, and experiences with ~~us~~ the authors. Four TAs were interviewed, in person or by email, and their feedback and experiences informed Dear Beta’s final design.

Both TA_1 and TA_3 adapted to the Dear Beta deployment by first asking each help-seeking student if they had already consulted Dear Beta. If they had not, TA_1 came back to them

Dear Beta: An Advice Column

Test Number **Find or Add Error** [Instructions](#) [Submit Feedback](#)

Request	File	Test Number	Hint
1	lab5/beta	29	Add a hint!
1	lab5/beta	33	<p>Add a hint!</p> <p>Upvote 2 If you're using muxes to choose Ra/Rb and 0, make sure the you're not using the same selector for both muxes.</p> <p>Upvote 2 Make sure that you're setting R31 to be 0</p> <p>Upvote 0 Make sure to check if BOTH Ra and Rb are R31.</p>
0	lab5/beta	38	Add a hint!
0	lab5/beta	43	<p>Add a hint!</p> <p>Upvote 0 Make sure your XP value is being set correctly.</p>

Figure 1-2: *Dear Beta* serves as a central repository of debugging advice for and by students, indexed by autograder errors. In this figure, there are three learnersourced hints, sorted by upvotes, for an autograder error on test ~~no.~~ 33 in the 'lab5/beta' checkoff file.

Request 4 **lab5/regfile** **2** **Submit**

Figure 1-3: After fixing a bug, students can add a hint for others, addressing what mistake prevented their own solution from passing this particular ~~verification~~ autograder test.

after visiting everyone else in the lab help queue. By then, they had often already resolved their problem with Dear Beta hints, and had a new bug they wanted help debugging.

Dear Beta was used as a debugging aid for both students and teachers. TA₂ described Dear Beta as a “starting point” for students, many of whom used it diligently during debugging. TA₂ appreciated that students who did ask for her help no longer said, “My Beta isn’t working. Tell me why.” Instead, they used Dear Beta as a starting point, to help them identify potential locations of a bug in many pages of code. Not just helpful for students, TA₃ was able to describe with specific examples how Dear Beta *helped him* help students quickly resolve common bugs.

TA₂ wondered if the extra hints were making it too easy to complete the lab, possibly letting students pass without understanding. TA₃ echoed this concern, but he made sure each student actually understood the Dear Beta hints whenever he personally guided them through the debugging process.

TAs identified both strengths and weakness in the Dear Beta design. TA₁ strongly supported Dear Beta’s existing design as a single scannable sorted list for quickly finding hints, rather than a purely search-based hint retrieval mechanism or the more general class forum. However, the affordances for contributing new hints in the initial prototype were not obvious and rarely visible on small screens. As a result, TA₂ was concerned that the level of student involvement in producing hints might be too low. The final Dear Beta design is more externally consistent with other participatory Q&A systems, has more salient buttons for contributing new hints, and a responsive design that accommodates screens as small as that of a cell phone.

TA₄ was absent during most of the Dear Beta deployment but still regularly recommended Dear Beta to students who asked for her help over email. A fifth TA declined to be interviewed. She felt that she had not interacted with Dear Beta enough.

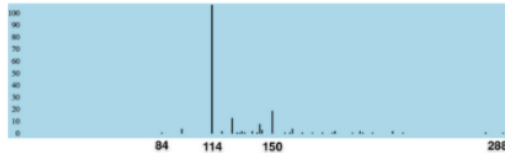
1.2.2 Dear Gamma

~~In order to~~ To learnersource optimization hints, ~~we caught students at a different stage in their learning process:~~ students participate in the mandatory Dear Gamma hint writing activity right after they ~~passed all verification~~ first pass all autograder tests for a particular digital circuit, the Full Adder. ~~Because students may have arrived at their solution without encountering any particular optimization obstacles, Dear Gamma uses the comparison workflow for learnersourcing rather than the self-reflection workflow.~~ The student has reached correctness so they are not in a position to think about optimality.

To collect Dear Gamma hints, students are given a pair of solutions and asked to give a hint to future students about how to improve from the less optimal solution to the more optimal solution. Students write hints for two such pairs of solutions. In each pair, one of the solutions is always their own. When the student's own solution is the better solution in the pair, then the student can hint at what the author of the less optimal solution might have missed. For example, the student might write, *Remember DeMorgan's Law: you could replace the 'OR' of 'ANDs' with a 'NAND' of 'NANDs.'* When the student's own solution is the poorer solution in the pair, they are challenged to first understand how the better solution uses fewer transistors, and then write a hint about the insight for a future student who authored a suboptimal solution like their own.

The comparison workflow was modified slightly, to accommodate the requirements of the course lecturer, who wanted to make sure that all students get a chance to consider both the most common and the most optimal solutions. The collection of previous student solutions in Figure ?? was also curated by the lecturer. If a student's solution is larger than the most common solution, they are not shown solutions larger than their own. Instead, they are asked to consider both the most common and the most optimal solutions, so they benefit from seeing both without doing extra work overall. Students with the most optimal solution only consider alternative solutions that are worse than theirs. Figure ?? shows an example of the page for a student with a 114-transistor solution. This was a required activity for all students during one semester of the course and 435 student-written hints were collected from approximately two hundred students who were enrolled in the course. The results of this deployment of the modified workflow is shown in Figure ?? as a Sankey diagram of the

Your design has a total of **114** transistors. For comparison, the graph below shows the number of designs submitted in a previous semester (y-axis) vs. their transistor count (x-axis).



Below we will ask you to compare your design with a few other designs submitted previously. First, for reference, here's your code:

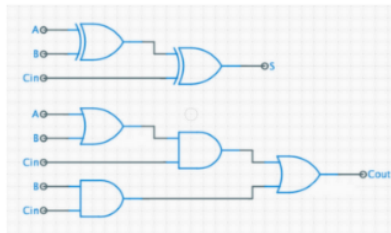
```
*nand
.subckt nand2 a b z
MPD1 z a 1 0 NENH sw=2 sl=1
MPD2 1 b 0 0 NENH sw=2 sl=1
MPU1 z a vdd vdd PENH sw=2 sl=1
MPU2 z b vdd vdd PENH sw=2 sl=1
.ends

*nor
.subckt nor2 a b z
```

Comparison #1:

If we used the design shown at the right for the FA module, a 3-bit adder would require **132** transistors, larger than your design by 18 transistors.

Imagine you're an Lab Assistant in a future semester of CompArch and a student submits a solution like the shown on the right. What advice would you give them on how to make their solution as good as yours?

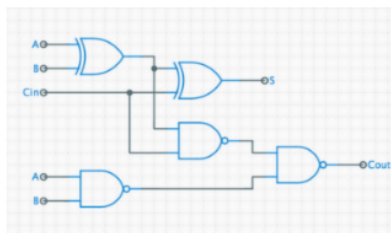


... enter your advice here

Comparison #2:

If we used the design shown at the right for the FA module, a 3-bit adder would require only **96** transistors, smaller than your design by 18 transistors.

Imagine you're an LA in a future semester of 6.004 and a student submits a solution like yours. What advice would you give them on how to make their solution as good as the one in the figure to the right?



... enter your advice here

Figure 1-4: Dear Gamma interface for a student with a solution containing 114 transistors. In the first comparison, they are asked to write a hint for a future student with a larger (less optimal) correct solution. In the second comparison, they are asked to write a hint for a future student with a solution similar to their own so that they may reach the smallest (most optimal) correct solution.

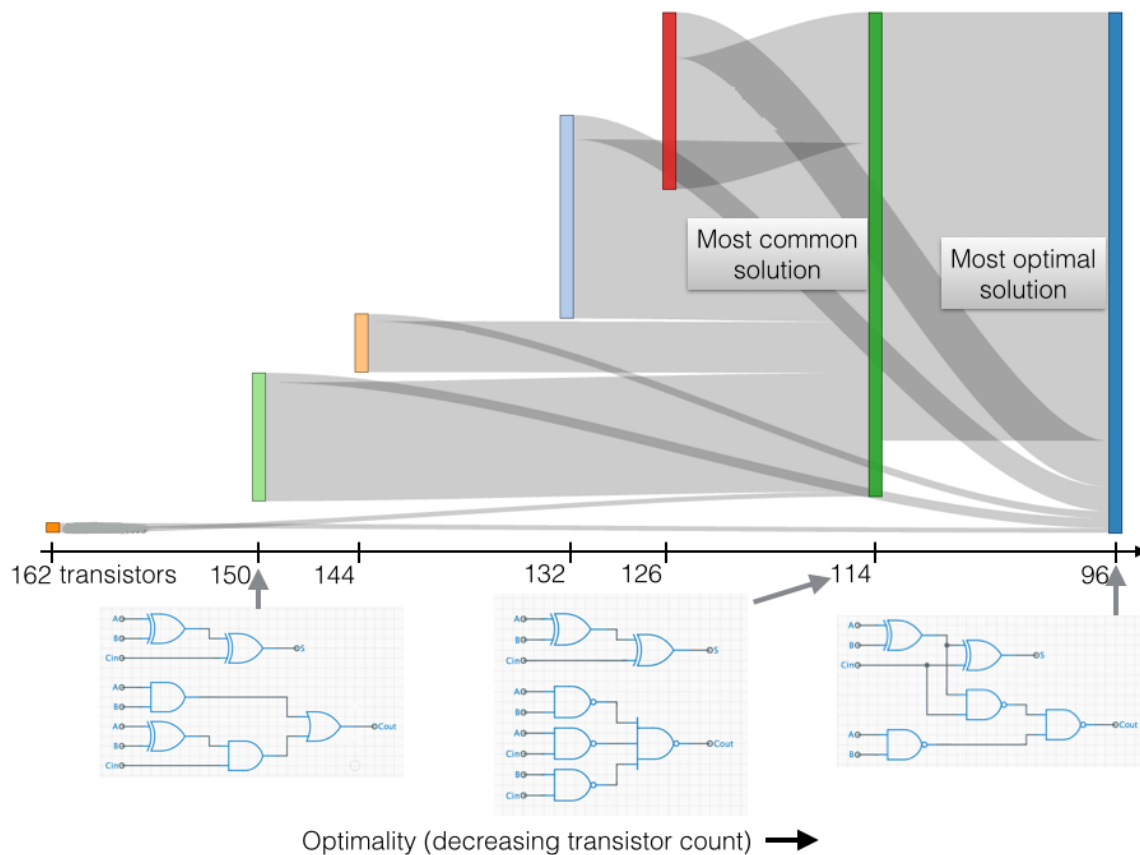


Figure 1-5: Sankey diagram of hints composed between types of correct solutions, binned by the number of transistors they contain. The optimal solution has only 21 gates and 96 transistors while the most common solution generated by students has 24 gates and 114 transistors.

distribution of solutions connected with student-written hints.

1.3 Evaluation

To evaluate the extent to which learnersourced hints can support problem solving, ~~we deployed~~ Dear Beta and Dear Gamma were deployed to the computer architecture class, which had an enrollment of more than two hundred students. Dear Beta was deployed for 6 weeks, ~~during which we~~, During that time, the system collected student-generated debugging hints ~~and~~, while the authors and course staff observed the simultaneous usage of those hints ~~in a real-world setting~~ by students working on class assignments. The Dear

Gamma optimization hint collection interface was released to students as a mandatory part of a particular lab. ~~We then~~ During a later offering of the same course, the authors conducted a lab study with nine students to understand how they solve a typical engineering problem using ~~these~~ the learnersourced optimization hints. The questions our evaluation sought to answer are: (1) *What are the characteristics of student-generated hints?* and (2) *Can students solve problems using those hints?*

1.3.1 Dear Beta

The Dear Beta website was released as a stand-alone additional resource for students one week prior to the due date for the final circuit design lab. Students were made aware of its existence through a class forum announcement and signs on chalkboards in the course computer lab. It was left up for the remainder of the semester for students to refer to, if completing work late. ~~We~~ The system tracked student logins and engagement with site features. An initial prototype of Dear Beta was also deployed for two consecutive semesters prior to this final system design and study, ~~as well~~.

1.3.2 Dear Gamma

Hint Succession and Categorization

While Dear Beta makes all hints available at all times, Dear Gamma is modeled on the hint-giving mechanism of an intelligent tutoring system. In prior work, sequences of hints have been posited to facilitate learning due to their similarity with sequences used in expert human tutoring, as well as their support of human memory processes [?]. Therefore, ~~we~~ further the authors decomposed the hints collected with Dear Gamma into the three kinds of hints that typically comprise a hint sequence: 1) *pointing hints* direct the student's attention to the location of error in case the student understood the general principle but did not know to apply it; 2) *teaching hints* explain why a better solution exists by stating the relevant

principle or concept; 3) *bottom-out hints* indicate concretely what the student should do [?].

Two researchers independently categorized the 435 collected Dear Gamma hints into six different categories: pure pointing hints (p), pointing and teaching hints (pt), pure teaching hints (t), teaching and bottom-out hints (tb), pure bottom-out hints (b), and hints that are irrelevant or clearly not helpful. They first independently labeled the first 30 hints. After discussing disagreements and iterating on their understanding of the hint categories, the coders then categorized the remaining 405 hints.

If one coder labeled a hint as a hybrid between two categories (i.e., teaching and pointing) while the other coder labeled it with only one category (i.e., pointing), ~~we assigned the hint~~ the hint was assigned to the pure category (i.e., pointing) that was in common between the two coders' labels. ~~If there was no shared category across the two coders~~ When the categories assigned by each coder did not overlap, the hint was discarded. ~~We also excluded the~~ The minority of hints (3.2%) ~~that were~~ labeled as irrelevant or unhelpful were excluded.

Lab Study

Nine out of the 226 current students in the computer architecture course participated in the study. These students were recruited through a course forum post. ~~Participants were given and offered \$30 for the study, which lasted one hour. We informed students that we were studying the effectiveness of participating in one hour-long session. The author of this thesis ran all sessions as the experimenter. While obtaining informed consent from each student, the experimenter explained that the study would assess the effectiveness of previously collected hints for optimizing circuits so that they use fewer transistors.~~

~~During the study, we presented the Dear Gamma hints, which were collected from, not the skills of the student. Students were not told that the hints were written by students in a previous semester, of the same class specifically for the Adder problem. Instead, the experimenter presented Dear Gamma hints as anonymous, potentially helpful messages. Three~~

Students began the study by reviewing their solution to the Adder problem, which was due earlier in the term. All recruited students had completed it. During the study, three batches of hints were shown in the order of pointing, teaching, and bottom-out, but randomly selected within each category. The ~~lab for which the Dear Gamma hints applied was due earlier in the semester and all recruited students had completed it. Students began the study by opening up their previously completed lab and reviewing their solution. The experimenter noted down~~ experimenter noted the number of transistors in their solution, and randomly selected five pointing-type hints for a solution of that size from the Dear Gamma collection. For example, if the student had 114 transistors in their solution, they received five hints previously generated by students who had written a hint to help improve a 114-transistor solution. Because hints may be of variable quality, the researcher presented hints in batches of five to increase the chances that one of them might be helpful.

The experimenter then asked the student to reduce the number of transistors in their solution. The experimenter explained that there were two more batches of five hints ready for them if they became stuck. These two batches were teaching hints and bottom-out hints. Students could consult outside resources like the course website and Google as well.

After receiving each batch of hints, participants answered the following 7-point Likert-scale questions about each hint (1: strongly disagree, 7: strongly agree): (1) *“This hint taught me something.”* (2) *“This hint helped me get to a more optimal circuit.”* and (3) *“I feel more confident that I could solve a similar problem in the future.”* ~~We placed these questions~~ These questions were asked immediately after each batch of hints to capture user perception of hints at the time they occurred. However, to avoid slowing down the problem-solving process, participants were asked to explain their answers in writing only after the study, in the post-study questionnaire. This rating process was repeated for the teaching hints and bottom-out hints, even if students were able to solve the problem without asking for these hints.

After the study, users completed a post-study questionnaire regarding their overall impressions. Because users were shown a batch of hints at a time, all of which were student-

generated, ~~in~~ the post-study questionnaire ~~we added~~ included additional Likert-scale items, “I was able to find the most helpful hints and ignore the rest” and “Many hints felt repetitive,” to understand whether users felt they could adequately ignore irrelevant hints.

1.3.3 Limitations

These studies do not have control groups, so ~~we cannot make~~ the results cannot support any claims about student learning. ~~We can only report~~ Dear Beta was deployed in a real classroom setting while students worked on a relevant class assignment, and the results contain qualitative and quantitative measures of student engagement with the system only. Some of those observed behaviors and opinions may be derived from the participants’ sense of novelty, rather than the underlying value of the system. ~~We deployed Dear Beta in a real classroom setting, and in the context of a real assignment, for the purpose of observing natural interaction with the system.~~

1.4 Results

1.4.1 Dear Beta Study

For the week prior to the lab assignment due date, the number of registered unique users in the Dear Beta system rose linearly from 20 to 166. It plateaued at 180 by one week after the lab was due. For comparison, the total number of students in the class was 226. 119 students logged in more than once and many students logged in repeatedly.

In the 9 days between the Dear Beta release date and the lab due date, users added 76 autograder errors and 57 hints as a response to those errors. Half of the errors received at least one hint. Seven errors received as many as three hints. Figure ?? shows user engagement with the system over time. As soon as the initial stock of hints were available, students began upvoting them.

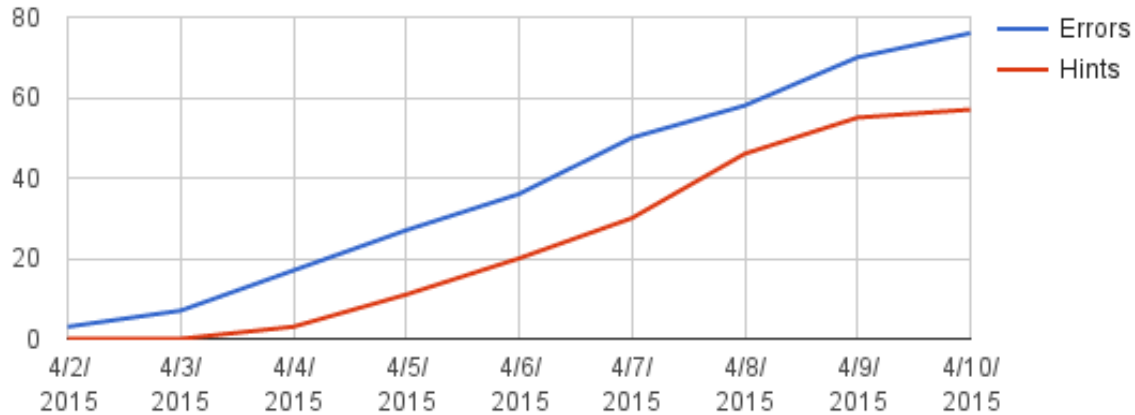


Figure 1-6: Between the Dear Beta release date (4/2) and the lab due date (4/10), autograder errors were consistently being entered into the system. The addition of hints followed close behind.

Users contributed 61 upvotes and 10 downvotes on the hints during the same period. The highest number of upvotes (10) was given to the hint “*When entering constants, 1#4 is 1111 and 1’4 is the 4-bit representation of 1.*” Remember that, while this is a teaching-type hint, it is provided as a targeted troubleshooting hint for students whose solution fails to pass a specific test case. The second most upvoted hint (5 upvotes) was “*Make sure your ASEL logic is correct - don’t allow the supervisor bit to be illegally set.*”

None of the hints appear to be incorrect, though this is difficult to verify, since the teachers do not have copies of the solutions from which these hints were generated. Even within a collection of hints for the same error, not all will be relevant to any particular solution.

1.4.2 Dear Gamma Study

With the Dear Gamma hints, six of the nine laboratory subjects were able to improve the optimality of their circuits within the hour that the study took place. Figure ?? illustrates the subjects’ revisions. One student only needed one set of pointing hints. Five students successfully revised their circuits after one set of pointing and one set of teaching hints. Four students received a set of final bottom-out hints as well. Three of those four students (S_2 , S_5 , and S_9) were still unable to successfully revise their circuits.

Hint Distribution Figure ?? is the Sankey diagram of the optimization hints collected by Dear Gamma. The number of hints between certain key transitions, such as between the most common and the most optimal solutions, was far greater because of the lecturer’s requests for pedagogically valuable hint prompts that introduced hint-writers to the common and optimal solutions.

The most common solution size was 114 transistors. Students with that common solution were randomly assigned to generate hints from one of the many different larger solutions down to theirs. These hints are pooled together with the hints written by students with solutions larger than 114 transistors who are seeing the common 114-transistor solution for the first time. Less than five percent of student solutions were the most optimal (96 transistors), but, at the request of the lecturer, every student was asked to consider that most optimal solution and write a hint for a fellow student on how to optimize their solution into that most optimal solution.

Students in the study were drawn from the same population as the hint-generating students, and all study subjects were offered the same number of hints (5 pointing hints, followed by 5 teaching hints, followed by 5 bottom-out hints) over the course of the hour-long session, regardless of the solution they started with.

Hint Types Table ?? shows the breakdown of hints by type, along with representative examples. The Cohen’s κ [?] inter-rater reliability was 0.54, which indicated that the two coders had moderate agreement across the six hint categories [?].

Hint Prompt Hint-authors interpreted the prompt to create a hint in different ways. Some addressed the hint-receiver directly (“*Keep in mind that...*”), while others addressed the teaching staff (“*I would mention [to the student]...*”). Some hint-authors did not directly write a hint, but instead wrote about how they would approach the situation of being a lab assistant for the hint-receiver: “*I think first I’d ask to make sure they knew what a NAND3 was, because I think a solution like this might come from not totally understanding how it works.*” Still others took a conversational approach, as if they were having an unfolding conversation with the hint-receiver. Interestingly, a number of hint-authors referred to “here”

Hint type	Count	(%)	Representative examples
Pointing (p)	62	14%	“You don’t have to keep S and Cout as two separate/independent CMOS gates.”
Pointing and teaching (pt)	81	19%	“Instead of making the S and Cout components individual, combine them together to save computation power.”
Teaching (t)	111	26%	“Instead of recalculating values, save computation results to save time!”
Teaching and bottom-out (tb)	19	4%	“Via application of demorgan’s theorem, $\text{NAND2}(\text{XOR } A \text{ } B) \text{ Cin}$ is equivalent to $\text{NAND3}(\text{NAND2 } A \text{ } B) \text{ Cin}$.”
Bottom-out (b)	78	18%	“Use the output of a xor b for one of the nand2 gates.”
Unhelpful or irrelevant	14	3%	“Use the hints provided by the lab, but try to improve on them.”
No coder agreement	70	16%	

Table 1.1: Breakdown of Dear Gamma hints by type. Students in the Dear Gamma lab study initially received 5 pointing hints (p), followed by 5 pure teaching hints (t), and finally 5 pure bottom-out hints (b), delivered whenever the student was stuck and asked for more help.

or “my circuit” in their hints, as if the hint-receiver would be looking at the Dear Gamma interface, with all its examples, rather than just the text generated by the hint-author. This particular assumption on the part of the hint-author was confusing for hint-receivers.

Optimization Issues S_5 was the only student who had a standard, optimizable solution, received hints, and had no insights about how to optimize the circuit within the allotted hour. S_1 , S_2 , and S_9 ’s forward progress was confounded by having near-optimal top-level architecture and very large (suboptimal) implementations of the underlying modules. Dear Gamma only shows hint-authors the top-level architecture, not the underlying gate implementations, for the alternative solutions they compare their own solutions to. They therefore found the hints, which were often about fixing high-level architecture, irrelevant and unhelpful. Even so, S_1 was still able to revisit the hints and correctly extract the lesson that only inverting gates should be used. As a result, S_1 successfully optimized their circuit.

While working through their optimizations and hints, S_6 was the one student who significantly deviated from the correct line of thought by removing all inverting gates. Optimal solutions have ~~only~~only inverting gates. As soon as S_6 saw that their transistor count had

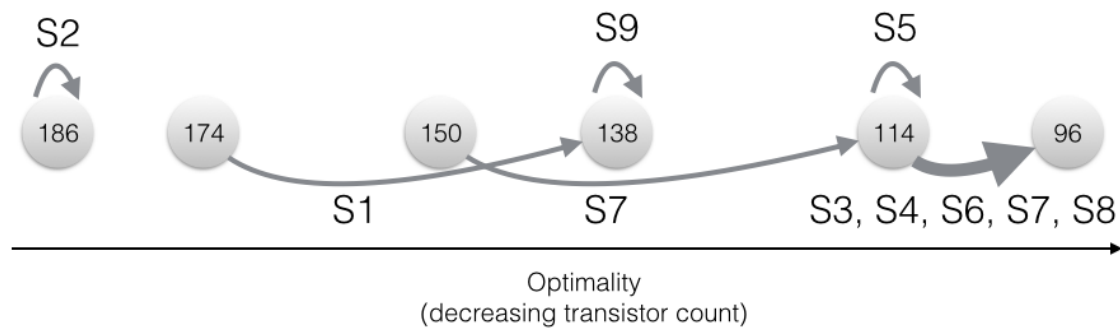


Figure 1-7: Six of the nine lab study subjects were able to improve the optimality of their circuits with the help of the Dear Gamma hints. Subject S7 was able to make two leaps—one to a common solution with 114 transistors and another from the common solution to the most optimal solution at 96 transistors.

increased rather than decreased, they revisited the hints, realized their mistake, and correctly optimized their circuit. None of the hints themselves were incorrect, though some were deemed irrelevant or unhelpful.

Hint Progression One student successfully optimized their solution from 150 transistors to the most common solution, 114 transistors, using only pointing and teaching hints. With some time left in the hour-long session, the student opted to optimize their circuit further. The experimenter gave the student one last set of hints, for the transition from 114 to the optimal 96-transistor solution. However, the experimenter did not restart the progression for this next transition. The student was given a set of bottom-out hints instead. Based on these hints, the student got the final optimization step without understanding, and appeared to feel cheated out of the satisfaction of figuring it out himself.

Student Reactions The six subjects without suboptimal gates agreed with the statement “Overall, these hints helped me get to a more optimal circuit” ($\mu=6$, $\sigma=1.1$ on a 7-point Likert scale). The remaining three subjects with suboptimal gates disagreed with the same statement ($\mu=2.6$, $\sigma=2.1$ on a 7-point Likert scale). Regardless of whether a subject’s solution included suboptimal gates, subjects on average agreed with the statement “Hints helped me think differently about the problem, even if they didn’t directly help me solve the problem” ($\mu=5.4$, $\sigma=1.6$).

Some subjects commented on the redundancy within each set of five hints of a particular type. This was sometimes expressed as a negative, as in *“These are all hinting at the same thing but I want new information,”* and sometimes expressed as a positive, as in *“Several hints are mentioning X.... I should look into it.”* One student told the experimenter that, while the individual hints were hard to understand, together they formed a clearer picture in her mind about what to do.

1.5 Discussion

In this section, ~~we first address~~ the research questions ~~the evaluation was intended to answer.~~ ~~We then explain that, of the design decisions made during the design of Dear Beta and Dear Gamma, the critical factors for success included~~ are revisited first. Then critical design decisions are reviewed, including prompt clarity, ~~the index-chosen index choice~~ for hints, ~~how alternative solutions are represented, and the use of~~ representation choice for alternative solutions, and hint progressions.

1.5.1 Answers to Research Questions

Our study sought to evaluate the characteristics of student-generated hints. ~~We can see from Table ??~~ Table ?? shows that students, without coaching, can naturally produce hints that point, teach, give a bottom-out direction, or provide some combination of those elements. However, the number of pointing hints labeled by *both* coders as purely pointing-type (22) was much smaller than the number of such hints in the teaching (75) and bottom-out (64) categories. Because students were not informed that their hint would be slotted into a progression, it is possible they may have felt that if they were going to give a future student one hint, it would need to be more substantial than just pointing to a particular location in the solution and hoping the hint-receiver would see the possibility of optimization.

Second, the studies sought to evaluate whether students can solve problems using these hints.

Both studies suggest that these student-written hints are helpful. The aggregate activity of students and teachers on Dear Beta indicate that the resource was populated with helpful hints. The Dear Gamma lab study was set up based on the observed sub-optimality of student circuits at the level of choosing and arranging gates. Students whose solutions were suboptimal in that anticipated way rated the hints as helpful. Students whose solutions were suboptimal in unanticipated ways, i.e., at the level of the gates themselves, were not well-served by the hints. Future optimization hint workflows will need both (1) an optimality metric that accounts for multiple common types of suboptimality and (2) a representation of solutions with an appropriate level of detail about the difference between any two solutions. Regardless, the Dear Gamma study suggests that students are helped by the hints when the optimality metric and representation are appropriate.

1.5.2 Lessons for Self-Reflection and Comparison Workflows

Prompt clarity appears to be critical for soliciting the highest possible quality of hints from students. In Dear Gamma, hint collection and delivery were separate processes. Some students misunderstood the prompt and wrote hints as if their audience was still the teacher, not a fellow student. Others did not understand that the hint-receiver would only see the text of their hint, not the diagrams it was based on. In Dear Beta, hint collection and delivery were all mediated through the same, constantly updated interface. The appropriate audience for a hint was clear. Future instantiations of the self-reflection and comparison workflows should clarify who the audience is for hint-authors, perhaps by displaying what students will see.

The selection of an index for hints in the self-reflection workflow matters. In Dear Beta, the choice of test file name and test number as an index for hints worked well for a class of hundreds of students. In a MOOC-sized course, the index may need to include an indicator that specifies how the test failed as well. Indices in future systems should have sufficient information to group related hints into clusters of a manageable size.

The success of the comparison workflow depends not only on the index for solutions, but also on how these solutions are represented in the workflow prompt for hints. In the comparison workflow, ~~we found that transistor counts sometimes did not account for~~ transistor counts did not always capture lower-level reasons for a suboptimal circuit, ~~resulting~~. This resulted in hints that were unhelpful for solutions with lower-level suboptimality. Students will not generate hints that account for what has been abstracted away in the representation of solutions in the hint prompt. Likewise, if the definition of optimality used to index solutions does not account for a certain kind of suboptimality, the hints generated will be unlikely to help future students with that kind of suboptimality.

Lastly, when students request hints as they did in the Dear Gamma lab study, conforming to the intelligent tutoring system model of providing progressively specific hints is recommended. To automatically create hint progressions in the future, ~~we could apply~~ machine learning methods ~~to estimate~~ could classify a given hint 's-type. Alternatively, ~~we could learnersource hint classification~~ hint type classification could be learnersourced.

1.5.3 Generalization

Although ~~we applied these workflows~~ these workflows were applied to computer architecture problems, the self-reflection and comparison workflows could be extended to other domains. The workflows can be most readily applied to solutions that can be objectively tested for satisfying a set of requirements, e.g., passing unit tests, or whose optimality can be objectively measured. In domains without objective test cases or definitions for optimality, it may be more difficult to establish indices for clustering hints. In these cases, students could be asked to write what challenge they overcame or select from a growing list, enabling others to search for those terms. The comparison workflow could be modified to simply pair students with solutions different than their own, letting them judge for themselves which they think is better, the alternative solution or their own, and write hints based on that judgment.

1.6 Conclusions

This paper enriches learnersourcing by shaping the design space for learnersourcing personalized hints, and presenting two workflows that engage students in hint creation while reflecting on their own work as well as that of peers. ~~We built~~ Dear Beta and Dear Gamma ; ~~which apply~~ are applications of these workflows to the creation of debugging and optimization hints, matching students to the appropriate hint creation task given their current progress. Results from our deployment study and subsequent lab study demonstrate the feasibility of these workflows, and indicate that student-generated hints are helpful to students. They also shed light on critical factors that may impact the quality of learnersourced hints, laying the groundwork for future systems in this area.