Peer conferences in calculus: The impact of systematic training

Daniel Lee Reinholz

Center for STEM Learning, University of Colorado, Boulder, Colorado, USA

393 UCB Boulder, Boulder, CO 80309, daniel.reinholz@colorado.edu

Daniel Reinholz is a Research Associate at the Center for STEM Learning at the University of Colorado, Boulder. He completed his PhD in mathematics and science education at the University of California, Berkeley, advised by Alan Schoenfeld. His research focuses on the role of social processes in promoting individual reflection.

Peer conferences in calculus: The impact of systematic training

This paper describes an intervention for improving the quality of peer assessment conferences in calculus. Although a body of work highlights the learning benefits of peer assessment, few papers have described the nature of student conversations during peer conferences/assessment in detail. This paper provides deeper insight into what those conversations actually look like, and shows the impact of systematic training on conferences. The study took place over two consecutive semesters of introductory college calculus, and analyses show that students had considerably improved conversations after training. The improved conversations consisted of much more on-topic talk and productive feedback; after training, students provided more feedback related to processes (communication and underlying reasoning) than products (correctness or incorrectness).

Keywords: calculus; feedback; reflection; peer assessment;

Introduction

This paper focuses on student learning through peer assessment in calculus. In particular, it describes the nature of conversations that took place between pairs of students as they analysed each other's work and provided feedback to one another. These peer conferences are an important part of peer assessment, because they allow for deeper social interactions than are possible when students only provide written feedback. This is crucial, because many theorists now emphasize the social nature of learning in mathematics (Lave 1996; Cobb et al. 1997; Boaler and Greeno 2000). These interactions allow students to engage in a number of useful activities, such as: explaining their ideas verbally to peers, co-constructing meaning around mathematical problems, and exchanging elaborated feedback. As a result of such activities, peer assessment can help students develop a number of skills, including: collaboration, communication, conceptual understanding, and problem solving (Falchikov 2005; Nicol, Thomson, and Breslin 2014).

Despite a growing body of research on the learning benefits of peer assessment, relatively little has been published on the nature of conferences that take place between students as they engage in peer assessment. Similarly, a body of work highlights the positive impact of training students to engage in peer assessment (e.g., Min 2006), but this work does not describe in depth how student feedback actually changes as a result of training. The present work seeks to strengthen the connection between these areas of research, by exploring the impact of training on the feedback provided during student conferences. This has implications both for researchers who aim to better understand how students support one another's learning through peer assessment, and instructors who aim to design better instructional systems for peer-supported learning.

The work reported here was conducted in the context of Peer-Assisted Reflection, an activity structure designed to help students learn to reflect on their own work by analysing the work of their peers (Reinholz 2015-b). To explore the nature of peer conferences, two semester-long phases of peer conferences in introductory college calculus were analysed. During Phase I, students engaged in the conferences with little training, while Phase II included a systematic training procedure. Thus, comparing these two phases of the study provides insight into the impact of training on student feedback. The following research questions are addressed:

1. What types of feedback did students provide during peer conferences?
2. What impact did systematic training have on the types of feedback provided?

Theoretical Framing

Assessment for Learning

Assessment for learning focuses on how to *evoke* information about learning and use it to *modify* teaching and learning activities (Black, Harrison, and Lee 2003). Broadly speaking, such formative assessment practices improve learning (e.g., Black and Wiliam 1998; Herman et al. 2014). More specifically, peer assessment is increasingly recognized as a core part of formative assessment (Black and Wiliam 2009). How peer assessment supports learning can be understood through the lens of the peer assessment learning cycle (Reinholz 2015-b). The assessment cycle focuses on six phases of practice: task engagement, peer analysis, feedback provision, feedback reception, peer conferencing, and revision. Each of these phases provides different learning opportunities. For instance, as students discuss their analyses with one another, they have opportunities to develop verbal explanation skills. This also supports conceptual understanding, given the close link between explanation and learning (Chi et al. 1994; Chi et al. 2001; Lombrozo 2006).

In addition to the above learning benefits, peer assessment supports the development of self-assessment skills (Reinholz 2015-b; Sadler 1989). Sadler (1989) argues that practical experience analysing various exemplars of a performance or solutions to a problem is a very powerful tool for learning how to make complex, qualitative judgments. Because there are so many potential dimensions along which to analyse complex activities, part of knowing how to make expert judgments is knowing which criteria are most relevant in a given situation. In other words, making such judgments requires developing an eye for what is "worth noticing" (Sadler 1989). By regularly engaging in peer assessment, students can help develop this sense of what is valued within a given discipline. This may be particularly valuable in mathematics, because argumentation and critique are such core disciplinary practices (Harel and Sowder 2007; Forman, Mccormick, and Donato 1997).

Feedback

Peer assessment allows students to *generate* and *receive* feedback. Generating feedback enables student to practise analytic skills that they can later apply to their own work, while receiving feedback helps students see their work from the perspective of their peers (Reinholz 2015-b). Both of these processes support self-assessment. The feedback that students generate and receive may be written or verbal, but in general, verbal feedback is more elaborated than written feedback and allows for students to clarify their feedback with one another. Beyond supporting self-assessment, receiving feedback helps an individual move from an existing state to a desired state of understanding or performance (Ramaprasad 1983). Once feedback is received, it is crucial that an individual actually uses the feedback to close the gap between actual and desired performance (Sadler 1989), which may help the learner better integrate new ways of thinking into their repertoire of practice.

Through a meta-analysis of feedback studies, Hattie and Timperley (2007) classified four types of feedback: (1) task, (2) processing of the task, (3) self-regulation, and (4) personal. Task feedback focuses on whether or not students correctly completed a task. Processing of the task focuses on the students' use of strategies and error detection methods, and how they make sense of what they are supposed to be doing. Self-regulation focuses on how students monitor, self-control, and direct their problem solving as they work through a task. Finally, person-focused feedback is related to the person who is engaging in the task, rather than the task itself (e.g., praise). Based on the review, not all feedback is equally useful, or even helpful. Feedback focused on processing of the task and self-regulation is the most beneficial to learning. In contrast, praise focuses individuals on themselves rather than on the task at hand (Mueller and Dweck 1998), which can actually inhibit learning. Similarly, grades can distract students from using elaborated feedback (Butler 1988).

For the present work, Hattie and Timperley's (2007) categories are condensed into three broad categories of feedback (in order of effectiveness): (1) process-focused, (2) product-focused, and (3) person-focused. In this categorization, processing of the task and self-regulation are combined into a single category, process-focused feedback. Conceptually, both of these categories focus on how students interact with the task, choose strategies, reflect upon their strategies, and make sense of what they are doing. These are all core parts of mathematical problem solving (Schoenfeld 1985). In this way, it makes sense to collapse the categories into one, which focuses on students' process of working through the task. Empirically, these two categories result in the highest learning gains. Moreover, in the specific context of peer feedback, there is some evidence that students are more likely to respond to comments asking them to explain in more depth (related to processes), as compared to feedback describing factual correctness (Walker 2015). Thus, the learning benefits of such process-focused feedback may even be amplified in a peer assessment context, as students may be more likely to incorporate it into their revisions.

Training

Given that not all feedback is equally useful, and that peer assessment is a novel activity for students, adequate training and scaffolding are required to help students engage (Smith, Cooper, and Lancaster 2002; Topping 2009). Training typically involves activities such as: working through assessment criteria with students, analysing work as a class, and teachers modelling assessment practices. Studies show that after training students are more likely to make useful revisions from the feedback that they receive, which is an indicator that they are receiving better feedback (Berg 1999; Min 2006). However, these studies do not describe the actual qualitative changes in feedback provided by students in depth, which is an emphasis of the present paper. Understanding these changes would support the design of instruction to improve feedback and provide further insight into how peer feedback supports learning.

Design

Peer-Assisted Reflection

The peer conferences studied in this paper took place in the context of the activity Peer-Assisted Reflection. Prior studies show that this activity has had a significant impact on student success; student pass rates during Phase I were improved by 13%, and with the addition of systematic training were improved by a greater extent, 23%, during Phase II (Reinholz 2015-a). Engaging in this activity also improved the quality of student explanations (Reinholz forthcoming).

As a part of the design, students engaged in a peer feedback cycle once each week with a challenging, in-depth homework problem. The cycle consisted of four activities: (1) students worked on the homework problem individually, (2) students completed a self-reflection on their work, (3) students came to class and exchanged feedback with a peer, and (4) students revised their work before turning in a final solution. Because students actually revised their solutions based on the feedback they received, they were able to close the feedback cycle (Sadler 1989). Students completed a total of 14 these challenging homework problems during the semester.

The present analyses focus on students' peer conferences. To begin their conferences, students traded their work with a peer, then silently reviewed their partner's work and provided written feedback for five minutes. This feedback was written on a peer feedback form that instructed students to provide feedback about both the communication and correctness of their partner's solutions (see Figure 1). After writing their feedback, students discussed their analyses with one another for five minutes. The primary purpose of the feedback form was to serve as a basis for these conferences; it was not expected that the written comments would be a major source of feedback for students because students could speak more completely in their conversations.

Figure 1. Prompts from the peer feedback form.

1. **Communication:** Give at least one suggestion to improve the communication of the solution.
2. **Correctness:** Note any errors you found.
3. **(Optional:)** What other feedback do you have? How else could the solution be improved?

Peer-Analysis Training (Phase II Only)

During Phase II, students received systematic training describing how to analyse student work and provide feedback. This activity took place once a week, immediately after students turned in their final solutions to the special homework problem. Students were given three sample solutions to one part of the problem that they had just completed; then they were asked to classify the solutions according to their quality and also to explain how the solutions could be improved. The activity was called "darts," because students had to identify the bull's-eye (best solution), on the board (mostly correct), and off the mark (incorrect) solutions.

To illustrate the darts activity, a sample student interaction is given. This interaction took place around the tenth Peer-Assisted Reflection task, which was focused on estimating the area of one's hand (see Figure 2). This task was introduced to students before they had formally studied Riemann sums. Students were intended to devise a method to estimate the area of their hand by breaking their complex hand shape into simpler shapes, calculating the area of the simple shapes, and then finding the sum of these estimates. This was approach evident in student work, as most students used squares, rectangles, triangles, and other basic geometric shapes to arrive at their estimates.

Figure 2. Peer-Assisted Reflection task 10: Hand Area

In this problem, you will trace the shape of your hand and approximate the area of the picture that you create. Your main tasks are to devise a method for approximating the area and to show that your approximation is very close to the actual area.

1. Put your hand flat on the grid provided (with fingers touching, no gaps) and trace the shape of the outline of your hand. Make sure that the shape you trace is a function (if not, erase the parts of the shape that would make it not a function).
2. Devise a method to approximate the area of the region inside the curve you have traced.
   Explain your method in detail, and explain why it should work. (Don't perform any calculations yet.)
3. Use the method you described above to approximate the area of the outline of your hand. (Show your work.)
4. Describe a method for estimating the error in your method of approximation. (Error is something you would like to make *small!* Thus an estimate for the error means being able to say the error is **less than** some value.)
5. Calculate an estimate for the error for your method.
6. Explain (in principle) how you could improve your method to make your estimate as accurate as one could want (i.e., minimize the error). (You do not actually have to perform the calculations, just explain what you would do.)

The corresponding darts training activity for this problem is given in Figure 3. The references to rectangles in the sample solutions are the simpler shapes that a student might use to estimate the area of their hand.

Figure 3. Peer-Assisted Reflection task 10 sample solutions for training

Prompt: Explain (in principle) how you could improve your method to make your estimate as accurate as one could want (i.e., minimize the error).

Sample 1: If I took a limit as the width of the rectangles approaches 0 (making the number of rectangles approach ∞), the difference in the area under the curve and the rectangles would approach 0.

Sample 2: You could use midpoints rather than endpoints and it will be more accurate because there will be less overlap.

Sample 3: If I had more rectangles there would be less overlap and the approximation would be better.

For the training activity, students spent a few minutes writing down their initial thoughts about each of these sample solutions before having a whole-class discussion. The following discussion highlights how students offered ideas about the second sample solution above (see lines 4 and 10) and how the instructor opened up space for students to share their ideas (lines 3, 9, 16, 18).

3. Instructor: What about number 2?
(Three students shake their heads no: Patrick, Colton, and Barry)
4. Patrick: In the lab we just did, we created that graph to show that midpoints aren't always more accurate.
…
8. Sue: Wouldn't midpoints be better?
9. Instructor: What do you think, would midpoints be better?
10. Barry: Would it even matter, because it says "as accurate as you would want," and you can only get so accurate with midpoints?
…
16. Instructor: What about last one? If I had more rectangles there would be less overlap, and the approximation would be better?
17. Dru: I said for two and three they are both on the board, because they both give an idea, but it's not fully developed.
18. Instructor: So what would you want them to add to this?

Of the students who responded, Patrick, drew on the calculator homework assignment that the class had recently completed to argue that midpoints are not always the most accurate method for approximation (line 4). In line 8, Sue asked for clarification about this suggestion. In response, Barry explained why the midpoint method is insufficient to produce arbitrary accuracy (line 10). After the students finished discussing the second solution, the instructor opened up space for the students to explain how to improve the third solution (lines 16 and 18). This interaction highlights that while the training activity was focused on communication, it created space for students to discuss mathematical concepts.

Method

Participants and Data Collection

Data were collected in two consecutive semesters of an introductory college calculus course at a research university. Each semester, approximately 400 students register for the course across 10 sections. Each section has 30 to 40 students enrolled on average, with a few larger sections having 50 to 90 students. All sections were taught with a common curriculum and common examinations. During each phase, there was a single experimental section that used Peer-Assisted Reflection and the nine other sections were used as comparisons. During the Phase I, a cooperating instructor (Michelle) taught the experimental section. Michelle had a PhD in mathematics education and nearly a decade of teaching experience. During Phase II, the researcher (Dan) taught the experimental section so that the systematic darts training method could be tested. In contrast to Michelle, the researcher was a graduate student with approximately three years of teaching experience.

A variety of data were collected, including: video observations, copies of student work and examinations, audio records of student conversations, student surveys, and interviews of students regarding their experiences with peer feedback activities (Reinholz 2015-a). From the larger corpus of data, this paper focuses on students' peer conferences and interviews. For each homework problem, a number of students were chosen randomly and their peer conferences were recorded. A total of 140 conversations were recorded (54 in Phase I, and 86 in Phase II). All of these conversations were transcribed for analysis. A total of 36 students were interviewed, with 14 from Phase I and 22 from Phase II. Interviews were selectively transcribed for the questions that were analysed for this paper. These questions were:

1. What type of feedback have you received from other students?
2. What type of feedback have you tried to give to other students?
3. To what extent has the "darts" activity supported or not supported you to be successful in Peer-Assisted Reflection, and in the course more generally? (Phase II only)

Note, for one student in Phase I, the feedback questions were not asked, because they were lost in the flow of the interview.

Analysis

Analyses focused on the amount and types of feedback provided by students during peer conferences. Three types of analysis were conducted to characterize the conversations: (1) content analysis of the entire corpus of conversations, (2) qualitative analysis of a subset of the conversations, and (3) student interviews in which students described the feedback they exchanged.

The goal of content analysis was to compare the amount and distribution of talk between students and the types of feedback exchanged during the two phases of study. The use of text mining techniques (the tm package in R) allowed for the large corpus of data (140 conversations) to be analysed objectively without the need to code student utterances. For text mining, categories of words were iteratively generated for the three categories of feedback previously identified: process, product, and person. To generate these lists, five conversations were randomly sampled from both Phase I and Phase II for a total of ten conversations. Reading through the transcripts, words in the conversations that related to these three feedback categories were identified; to expand the lists, an online thesaurus was used to find synonyms for all of these words. Text mining was then run on all 140 conversations with these elaborated lists, with all words that did not appear at least twice dropped from the lists. To prepare the transcripts for text mining, all punctuation and excess spaces were removed, and words were reduced to their roots (e.g., communic* for communicate, communication).

To provide a qualitative illustration of the quantitative results, student conversations for the tenth Peer-Assisted Reflection task were analysed. This task was chosen because it was late enough in the semester that differences in student conversations across Phase I and Phase II would be noticeable. Also, student explanations for this problem have been analysed in depth (Reinholz forthcoming), so it is possible to compare the quality of conversations to the quality of student work. Six student conversations from Phase I and seven conversations from Phase II were recorded for this task. Of these conversations, a random number generator was used to choose two conversations from each phase.

Finally, student interviews were analysed for the types of feedback that students said that they exchanged. This analysis was done to check for consistency between the actual content of student conversations and students' perceptions of their conversations. The training question was analysed to determine if students perceived the training as useful, regardless of actual learning benefits. Once all interviews were transcribed, student responses were randomized so that it was not possible to identify which phase the responses came from except for the question related to training. The researcher then coded the responses for the types of feedback exchanged. This allowed for student responses from Phases I and II to be compared without bias.

Results

The results are organized by the three categories described above.

Content Analysis

The average conversation length during Phase II was 635 words (SD = 252 words), while it was only 351 words for Phase I (SD = 173 words); both distributions were approximately normal. Conversations were also analysed for distribution of talk between students, by measuring the percentage of talk (in words) contributed by the student who spoke less. During Phase II, the average talk contributed by the quieter student was 34.5% of the conversation (SD = 11%), and it was 36% during Phase I (SD = 11%). Feedback was classified as focusing on: (1) processes, (2) products, and (3) persons, which represents a hierarchy from (1) to (3) of how useful the feedback is for learning (Hattie and Timperley 2007).

Process-Focused Feedback

To investigate process-focused feedback, conferences were analysed for the use of question words (Table 1). During conferences, students asked questions to clarify their partner's work. Because the written work generally made it clear which answer the students had arrived at (product-focused), these questions were used to elicit more information about how the answer was arrived at or what the student's underlying thinking was. For Phase I, students used an average of one question word per 56 words; Phase II students used an average of one question word per 55 words. Thus, even though the conversations were much longer during Phase II, the density of questions did not diminish. This meant Phase II students used an average of 11.5 question words per conversation, compared to 6.3 question words per conversation in Phase I.

Table 1. Frequency of question words

	Phase I (N = 54)	Phase II (N = 86)
Who	101	223
What	142	411
Where	54	129
When	23	120
Why	2	9
How	17	99
Total	339	991

Process-focused feedback was further explored by analysing student talk related to communication. When focusing on how ideas were conveyed to a peer, students generally had to examine their peers' underlying thought processes and make them more explicit (as was illustrated in the darts training example above). Table 2 contains the usage of communication-focused words. For Phase I, there were an average of 4.2 communication-focused words per conversation. For Phase II there were 7.7, nearly twice as many. This is consistent with the use of question words.

Table 2. Frequency of communication-focused words

	Phase I (N = 54)	Phase II (N = 86)
Communic*	4	17
Explain	48	67
Find	19	113
Mean	26	71
Read	8	27
Said	35	92
Say	37	143
Talk	26	52
Tell	7	38
Understand	15	43
Total	225	663

Product-Focused Feedback

Table 3 provides the frequency of product-focused words. Phase I students used an average of one product-focused word per 234 words, while Phase II students used an average of one product-focused word per 265 words; students in both phases spent a relatively similar amount of their talk on correctness or incorrectness. Given that Phase II students spent more time talking about processes, this meant that Phase I students focused a greater proportion of their feedback on products.

Table 3. Frequency of product-focused words

	Phase I (N = 54)	Phase II (N = 86)
Right	68	152
Wrong	11	28
Correct	2	14
Incorrect	0	2
Total	81	196

Person-Focused Feedback

During Phase I, students used an average of one person-focused word per every 68 words, while the average was one person-focused word per 120 words during Phase II (see Table 4). This indicates that conversations from Phase I had a greater emphasis on people, rather than the problems, than those during Phase II. Students essentially always used praise words; critical words such as “bad” were absent.

Table 4. Frequency of Person-Focused (Praise) Words

	Phase I (N = 54)	Phase II (N = 86)
Like	242	351
Love	0	5
Good	33	72
Great	3	4
Nice	2	11
Cool	2	14
Total	282	457

In summary, Phase I students used a greater proportion of their talk giving product-focused and person-focused feedback than during Phase II. In contrast, Phase II students provided a greater quantity of process-focused feedback during their conversations.

Qualitative Examples

To contextualize the quantitative analyses above, four examples of student conversations are analysed below. This sampling of conversations is generally consistent with the analysis above, showing that Phase II conversations were longer and more focused on the process of solving mathematical problems.

Phase I

The first of the sampled conversations (Adam and Beth) consisted of only 12 turns of talk. This conversation was mostly superficial, focusing only on two ideas: (1) how to pronounce Riemann, and (2) the trapezoid rule. However, the trapezoid rule was not discussed in any depth; Adam simply noted “you know how you have a left and a right sum, so the trapezoid just does this.” The two students did not provide any feedback on each other’s communication or correctness.

The second sampled conversation was of greater depth than the first, consisting of 21 turns of talk (Cody and Devin). At the beginning of the conversation, Cody told Devin that “I didn't know exactly what you were trying to do,” but did not provide any feedback on how Devin could improve his explanations. This was the only discussion that the students had about communication. Adam’s comment “but overall, you did a good job” was an example of the types of praise students provided to each other.

Although they did not focus on explanations, the students did talk about the mathematics in the problem. For instance, Devin suggested how Cody could improve his work (line 8), and then discussed his own method in detail (line 10):

8. D: start like, at 0, at the bottom of your hand, and just go all the way up and measure to the top. Do the lines one by one. That way you don't have to deal with the negative in your Riemann sum. When you draw the rectangle…
9. C: I liked that part of your graph.
10. D: then, the way I did my area, you could have a missing triangle here from the total area, the base times height of the triangle. You have a triangle here and a triangle here. I just took the area of those, and then added them, and subtracted them from the total area, and looked at that difference between the area I found and the Riemann sum was the error.

While Devin provided suggestions for Cody (lines 8 and 10), all of his feedback involved simply telling him the answers, rather than helping him arrive at the answers on his own. Later in the conversation (lines 11-21) students discussed their ideas for finding the error, but still did not ask any questions to further elicit one another’s ideas. Consistent with the quantitative findings above, students had relatively short conversations that did not focus much on mathematical processes.

Phase II

The sampled conversations from Phase II were much longer than those sampled during Phase I; the first conversation had 45 turns and the second had 39. Moreover, the students remained on topic throughout the entire conversations, unlike Adam and Beth's very superficial conversation.

Consistent with the quantitative analyses, these students spent more time focusing on communication. For instance, Eduardo (line 13) and Frank (lines 2 and 8) both provided direct feedback about communication:

2. F: Yours looks fine. The only thing I didn't quite understand was your method of finding the incomplete squares. How exactly did you do that?

…

7. E: So should I explain more about the incomplete?
8. F: Probably a little bit. That's the only, pretty much, like, help that I...or, like, um, what's it called, feedback for the communication.

…

13. E: Okay. For you communication, how you mentioned you would be over-estimating by using the right side. Thought that was really good to throw in there. You also mentioned the left side, so maybe calculate both and do the average. And then I saw you did went through and calculate your error, but there's no explanation for it.

This type of discussion was typical during Phase II, but very rare during Phase I. In the other sampled conversation, Hassan provided feedback about communication (line 8), and then later in the conversation George provided feedback (line 19) about Hassan's communication:

8. H: For communication, all I saw was explain how you got the integral. And for correctness, I mean, I'm sure you got a good estimate from that but, like, what I would do is split it in to max and min for each interval so you get a good range that your estimate is in. Instead of just, like

…

19. G: Okay. Yours, your instructions were really good. One thing is, you basically talked about the right hand and left hand endpoints like Riemann sums, but you don't actually say that's what it is. So you could say this is a Riemann sum.

After George provided communication-focused feedback, the students had a deeper conversation about the underlying reasoning of the solution (lines 20-22).

20. H: I mean, I kind of did that. But it's not left hand and right hand. I just did the max and min. Cause, like, if you look at this, my max for this one is not necessarily the left hand one. It's right there. Like, the middle. It's just the max on that interval.
21. G: Oh
22. H: And my min isn't necessarily, even though it is, it's not always necessarily the right hand one. It's just the minimum of that interval. Just so I could get a good

> range. An accurate range. So I did say the Riemann sum, well I didn't say it here. I was going to say it but that's not really the Riemann sum. It's like, a little bit different

This example illustrates how a focus on communication often pushed students to be more explicit about their underlying reasoning.

Interviews

The final component of analyses focused on student perceptions of the feedback exchanged. Given that perception of a phenomenon can be just as important as what actually takes place (e.g., Bandura 1997), these analyses were aimed to determine if students perceived their experiences consistently with their actual conversations. As the analyses below highlight, student perceptions were consistent with the differences above; Phase II students focused much more on processes than Phase I students.

Feedback Received

Phase II students indicated that they received much more process-focused feedback than Phase I students (see Table 5).

Table 5. Types of feedback received

	Process-Focused		Product-Focused
	Underlying Reasoning	Communication	Correctness
Phase I (N = 13)	2 (15%)	5 (38%)	8 (62%)
Phase II (N = 22)	4 (18%)	14 (64%)	15 (68%)

When students discussed receiving process-focused feedback, they made statements such as:

- "It's not like they are giving me an answer, they are giving me an idea where to start." (Phase I)

- "You didn't do this portion correctly, and here's the reasoning why it's not correct." (Phase II)
- "They'll give me an idea. So I'm like, this is how I'll do it tonight, now that I have an idea." (Phase II)

These quotes indicate that students received feedback on how to approach the problem, not simply what the right or wrong answer was.

Students rarely discussed receiving person-focused feedback. When they did, it was generally in the context of saying that they received superficial or unhelpful feedback. During Phase I, 4 out of 13 (31%) of students indicated that they received superficial or unhelpful feedback from their peers. In contrast, only 1/22 (4.5%) of students during Phase II indicated that they received feedback of this sort. These five responses are given:

- "Most of the time people don't really go that deep into it. They are like "yeah, good job, or no, the answer is this." (Phase I)
- "Usually they say everything looks right, so I don't really get that much back from it." (Phase I)
- "I've gotten a couple of good jobs, and I'm like, all right, cool, even though I've done the problem wrong." (Phase I)
- "One student, she said…I'm just trying to come up with something to put on this paper for you so you can get points. And I totally understand that, because that's what I do." (Phase I)
- "Most of the time, to be honest, it's pretty vague. I can see that I am pretty vague with people, and they are pretty vague with me. Unless someone has done something completely wrong, and you can tell that it is completely wrong, it's really hard for us students to say something is wrong." (Phase II)

As the examples illustrate, students indicated that they did not find praise to be particularly helpful (the first three quotes). The other two instances of superficial feedback did not focus on praise, but a general lack of benefit from the process.

Feedback Given

Consistent with feedback received, many more Phase II students talked about giving process-focused feedback, compared to Phase I students who were more likely to focus on products (see Table 6).

Table 6. Types of feedback given

	Process-Focused		Product-Focused
	Underlying Reasoning	Communication	Correctness
Phase I (N = 13)	4 (31%)	3 (23%)	10 (77%)
Phase II (N = 22)	6 (27%)	19 (86%)	13 (59%)

All but three of the students in Phase II emphasized a focus on communication. During both phases, students made comments such as:

- "Usually with the problems, I think most people get the problems right, so you don't have to spend a whole lot of time on the whole correctness stuff. So I just try to say if there's anything they could do better about how they explained their solution." (Phase I)
- "Basic communication, if I can't understand what they were trying to say very well or what they did, I will say, you should try to explain this better and tell them explicitly what I was confused about in their explanation or problem." (Phase II)
- "I try to give a lot of feedback on explaining it more. A lot of people, I've seen some where people just write the answers down. Or they skip many many many steps to the point where you can't follow it. I think being able to follow how they did it, even without writing, is important to do." (Phase II)

Compared to feedback received, more students reported focusing on underlying reasoning in the feedback they provided. When referring to underlying reasoning, students made comments such as:

- "When we talk about it, I'll ask why does it do this, and where did you get this from?" (Phase I)
- "I try not to give them answer, but how to do the problem…I guess hints at the right answer instead of giving it straight out." (Phase II)

Students did not discuss providing person-focused feedback. This lack of focus is consistent with the above finding that students generally described such feedback as superficial and unhelpful; in the interviews all students indicated that they at least tried to help their peers to the best of their abilities.

Darts Training

When asked if they found darts to be helpful or not, 18 students indicated it was helpful, 2 said it was not helpful, and 2 said they were not sure whether or not it was helpful. The four students who did not indicate that darts was helpful were unable to provide elaborated reasons, making statements such as “I don’t know; it doesn’t really help me.” When students talked about why it was helpful, they noted that it was useful to see different ways of wording explanations and that it helped them see the difference between a good (on the board) and great (bull’s-eye) solution. For instance:

- “For good examples of justifying your solution and really bad examples it is helpful. You can look back at your own work more objectively and notice, that totally sucks, I should fix it.”
- “It definitely helps you look at how other people explain things. So if someone else has a really good explanation, you're like oh yeah that totally makes sense. That's how it needs to be explained. So it helps you figure out better ways to explain the problems. If you look at someone's explanation and see that you have no idea what's going on, then you realize you have to include that in your explanations, and to add this, or whatever.”
- “Something that I really took out of the darts was that whole difference between an okay answer and an excellent answer. So I'll be doing my [Peer-Assisted Reflection homework], and I'll say, right now what I have, will that give me on the board, or a bull’s-eye. So what's the difference between that? I'm trying to get a bull’s-eye. And that's something I really learned from it.”

These findings suggest that the training activities helped students develop a new framework for thinking about the quality of mathematical solutions. More than just helping students provide better feedback, training made students more aware of the nuances in the quality of work, which was a skill that could be transferred over to assessment of their own work. In other words, it made students more aware of what they should focus on in analysing the quality of work (Sadler 1989).

Conclusion

This paper focused on the nature of peer conferences in calculus. The conferences took place in the context of Peer-Assisted Reflection, an activity structure that has been shown to have a significant impact on student success. Student conferences were compared over two phases of study, one that included systematic training on feedback and analysis of mathematical solutions, while the other did not.

The first research question focused on the nature of student conversations. During both phases of study, students used the conferences to provide multiple types of feedback, focusing on processes, correctness, and praise. Students were explicitly prompted to provide feedback on communication (related to processes) and correctness, but the affective nature of discussions (related to praise) was not emphasized in any of the class sessions. Nevertheless, students still evidently found it important to provide praise-focused feedback. During both phases, the distribution of talk was relatively even between students, with the quietest students still speaking about 35% of the time on average. Even without training, Phase I students were able to use their peer conferences in productive ways to support learning.

During Phase II, a training procedure was implemented to help students provide better feedback to one another. Addressing research question 2, training had a considerable impact: the average length of student conversations nearly doubled during Phase II, without a reduction in the quality of talk. Thus, Phase II students spent much more time discussing mathematics than those during Phase I. Given the importance of explanation and discussion in learning mathematics, this meant that Phase II students were better able to avail of the potential learning opportunities of peer conferencing.

The nature of student talk also shifted during Phase II. Students spent more time focused on processes rather than products or praise. Given that process-focused feedback is the most useful type of feedback for supporting learning, this meant that students made better use of the peer conferences. In this way, the emphasis of student conferences became more aligned with the training discussions during class, which were focused on the underlying thought processes of student work. By engaging in such conversations on a regular basis, students learned what was important to focus on in this particular disciplinary context. Consistent with the greater learning gains found during Phase II (Reinholz 2015-a), increasing process-focused feedback appeared to improve conceptual understanding. This is a productive area for future study.

While the present study highlights the impact of training on student conferences, the actual quality of different types of feedback was not analysed. Given the greater length of conferences and increased emphasis on processes, one might expect that the way in which Phase II students provided feedback about processes might shift as well, but it is not possible to determine based on the present analyses. This is an area for future study. Future studies might also focus on the written feedback that students provided, which were not analysed here. Given that Peer-Assisted Reflection is a learning activity that has also been used productively in other learning contexts (e.g., physics; Dounas-Frazer & Reinholz forthcoming), it would also be worthwhile to study peer conferences in those contexts and compare the discussions to those that took place around mathematical problem solving.

Acknowledgments

The author thanks Elissa Sato and Lina Haldar for their helpful feedback on an earlier version of this manuscript. The research reported here was supported in part by the

Institute of Education Sciences pre-doctoral training grant R305B090026 to the University of California, Berkeley. The opinions expressed are those of the author and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

References

Bandura, A. 1997. *Self-Efficacy: The Exercise of Control*. New York: Freeman.

Berg, E. C. 1999. "The Effects of Trained Peer Response on ESL Students' Revision Types and Writing Quality." *Journal of Second Language Writing* 8 (3): 215–41. doi:10.1016/S1060-3743(99)80115-5.

Black, P., C. Harrison, and C. Lee. 2003. *Assessment for Learning: Putting It into Practice*. Berkshire, England: Open University Press.

Black, P., and D. Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education* 5 (1): 7–74.

———. 2009. "Developing the Theory of Formative Assessment." *Educational Assessment, Evaluation and Accountability* 21 (1): 5–31.

Boaler, J., and J. G Greeno. 2000. "Identity, Agency, and Knowing in Mathematics Worlds." In *Multiple Perspectives on Mathematics Teaching and Learning*, edited by J. Boaler, 171–200. Westport, CT: Ablex Publishing.

Butler, R. 1988. "Enhancing and Undermining Intrinsic Motivation: The Effects of Task-Involving and Ego-Involving Evaluation on Interest and Performance." *British Journal of Educational Psychology* 58 (1): 1–14.

Chi, M.T.H., N. De Leeuw, M.H. Chiu, and C. LaVancher. 1994. "Eliciting Self-Explanations Improves Understanding." *Cognitive Science* 18 (3): 439–77. doi:10.1016/0364-0213(94)90016-7.

Chi, M.T.H., S.A. Siler, H. Jeong, T. Yamauchi, and R.G. Hausmann. 2001. "Learning from Human Tutoring." *Cognitive Science* 25 (4): 471–533.

Cobb, P., A. Boufi, K. McClain, and J. Whitenack. 1997. "Reflective Discourse and Collective Reflection." *Journal for Research in Mathematics Education* 28 (3): 258–77.

Dounas-Frazer, D. R., & Reinholz, D. L. forthcoming. *Attending to lifelong learning skills through guided reflection in a physics class*. Retrieved from http://arxiv.org/abs/1503.01788

Falchikov, N. 2005. *Improving Assessment through Student Involvement: Practical Solutions for Aiding Learning in Higher and Further Education*. New York, NY: Routledge.

Forman, E.A., D.E. Mccormick, and R. Donato. 1997. "Learning What Counts as a Mathematical Explanation." *Linguistics and Education* 9 (4): 313–39.

Harel, G., and L. Sowder. 2007. "Toward Comprehensive Perspectives on the Learning and Teaching of Proof." *Second Handbook of Research on Mathematics Teaching and Learning* 2: 805–42.

Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112. doi:10.3102/003465430298487.

Herman, J., S. Epstein, S. Leon, D. L. T. Matrundola, S. Reber, and K. Choi. 2014. "Implementation and Effects of LDC and MDC in Kentucky Districts."

CRESST Policy Brief No. 13. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
Lave, J. 1996. "Teaching as Learning, in Practice." *Mind, Culture, & Activity* 3 (3): 149–64.
Lombrozo, T. 2006. "The Structure and Function of Explanations." *Trends in Cognitive Sciences* 10 (10): 464–70. doi:doi:10.1016/j.tics.2006.08.004.
Min, H. 2006. "The Effects of Trained Peer Review on EFL Students' Revision Types and Writing Quality." *Journal of Second Language Writing* 15 (2): 118–41.
Mueller, C. M., and C. S. Dweck. 1998. "Praise for Intelligence Can Undermine Children's Motivation and Performance." *Journal of Personality and Social Psychology* 75 (1): 33–52.
Nicol, D., A. Thomson, and C. Breslin. 2014. "Rethinking Feedback Practices in Higher Education: A Peer Review Perspective." *Assessment & Evaluation in Higher Education* 39 (1): 102–22. doi:10.1080/02602938.2013.795518.
Ramaprasad, A. 1983. "On the Definition of Feedback." *Behavioral Science* 28 (1): 4–13. doi:10.1002/bs.3830280103.
Reinholz, D. L. (forthcoming). *Capturing changes in calculus explanations supported by peer review.*
Reinholz, D. L. 2015a. Peer-Assisted Reflection: A design-based intervention for improving success in calculus. *International Journal of Research in Undergraduate Mathematics Education*. http://doi.org/10.1007/s40753-015-0005-y
Reinholz, D. L. 2015b. The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*: 1–15. http://doi.org/10.1080/02602938.2015.1008982
Sadler, D.R. 1989. "Formative Assessment and the Design of Instructional Systems." *Instructional Science* 18 (2): 119–44. doi:10.1007/BF00117714.
Schoenfeld, A. H. 1985. *Mathematical Problem Solving*. New York: Academy Press.
Smith, H., A. Cooper, and L. Lancaster. 2002. "Improving the Quality of Undergraduate Peer Assessment: A Case for Student and Staff Development." *Innovations in Education and Teaching International* 39 (1): 71–81.
Topping, K. J. 2009. "Peer Assessment." *Theory Into Practice* 48 (1): 20–27.
Walker, M. 2015. "The Quality of Written Peer Feedback on Undergraduates' Draft Answers to an Assignment, and the Use Made of the Feedback." *Assessment & Evaluation in Higher Education* 40 (2): 232–47. doi:10.1080/02602938.2014.898737.