International Journal for the Scholarship of Teaching and Learning

Volume 12 | Number 2 Article 10

July 2018

Three Approaches to Focusing Peer Feedback

Daniel L. Reinholz
San Diego State University, daniel.reinholz@sdsu.edu

Recommended Citation
Reinholz, Daniel L. (2018) "Three Approaches to Focusing Peer Feedback," *International Journal for the Scholarship of Teaching and Learning*: Vol. 12: No. 2, Article 10.
Available at: https://doi.org/10.20429/ijsotl.2018.120210

Three Approaches to Focusing Peer Feedback

Abstract

Peer assessment has great potential to improve student learning. However, assessment is not an everyday activity for students, and thus providing appropriate guidance to students is a key component of creating a successful peer assessment experience. This paper explores how to structure peer feedback in the guided process Peer-Assisted Reflection (PAR), by comparing the artifacts and practices associated with three different iterations of PAR in undergraduate calculus. The iterations are referred to as the Questions, Critique, and Balanced approaches. Through a detailed analysis of this design-based research project, new insights are generated about how particular artifacts shape the feedback provided by students. In particular, students in the Balanced approach provided more succinct feedback across a greater variety of categories. In contrast, the Questions and Critique approaches had longer, narrative feedback, and it was focused on few categories.

Keywords

peer assessment, feedback, design-based research, mathematics

Cover Page Footnote

The author thanks Jason Siefken for creating the excellent graphics for the Balanced feedback form.

Three Approaches to Focusing Peer Feedback

Daniel L. Reinholz

San Diego State University

(Received 31 March 2017; Accepted 23 January 2018)

Peer assessment has great potential to improve student learning. However, assessment is not an everyday activity for students, and thus providing appropriate guidance to students is a key component of creating a successful peer assessment experience. This paper explores how to structure peer feedback in the guided process Peer-Assisted Reflection (PAR), by comparing the artifacts and practices associated with three different iterations of PAR in undergraduate calculus. The iterations are referred to as the Questions, Critique, and Balanced approaches. Through a detailed analysis of this design-based research project, new insights are generated about how particular artifacts shape the feedback provided by students. In particular, students in the Balanced approach provided more succinct feedback across a greater variety of categories. In contrast, the Questions and Critique approaches had longer, narrative feedback, and it was focused on few categories.

INTRODUCTION

Feedback is a fundamental part of learning (e.g., Hattie & Timperley, 2007; Shute, 2008). Yet, providing meaningful feedback poses major challenges for today's educators. Especially in large lecture courses, logistical constraints make it difficult to regularly provide in-depth feedback to students.

To address this challenge, many educators implement peer assessment in their courses (e.g., Dochy, Segers, & Sluijsmans, 1999; Topping, 1998). Peer assessment activities involve students providing feedback to one another about their work. When peer assessment is used in this way, for the purpose of learning, it can be considered a formative assessment (Black & Wiliam, 1998; Sadler, 1989). This contrasts summative assessments, or peer grading, which is used for the purposes of efficiently scoring student work. Formative peer assessments can have a variety of learning benefits for students, such as improved disciplinary understanding, self-assessment ability, and communication skills (Reinholz, 2015c).

How should such activities be organized? While assessment is part and parcel of being a teacher, it is something that students themselves rarely engage in as a part of the learning process. Thus, students need support. Research shows that when students are taught how to provide feedback to one another, the quality of their feedback improves (e.g., Min, 2006; Reinholz, 2015a). Similarly, rubrics can help guide students through the feedback process (Andrade & Valtcheva, 2009; Panadero & Romero, 2014). Yet, the critical question remains: how should these activities be organized? How do particular tools shape the feedback students give? This paper explores such questions.

To explore these questions, this paper draws from three iterations of a peer assessment activity in mathematics, revised using a design-based approach (Cobb, Confrey, Disessa, Lehrer, & Schauble, 2003). In each iteration, the feedback form and surrounding activities were revised, while instructional methods remained similar. Thus, by comparing student feedback across these various iterations of the activity, it is possible to gain insight into how students used different versions of the tool. Ultimately, the objective of this paper is to provide insight into how different artifacts used in this feedback activity influenced the types of feedback students provided. This will support instructors to thoughtfully consider how they develop artifacts to support student learning.

THEORETICAL FRAMING

Formative assessment is a highly-valued set of instructional practices, in mathematics (NCTM, 2000; Schoenfeld, 2014), and across disciplines (e.g., Black & Wiliam, 1998; Young & Kim, 2010). Such assessments concern how to *elicit* information about student thinking and *use* that information to improve the teaching and learning process (Black & Wiliam, 2009). When assessment is used formatively like this, it generally leads to better student outcomes (e.g., Black & Wiliam, 1998; Herman et al., 2014).

In this paper, peer assessment is defined as a particular type of formative assessment through which students analyze the work of their peers and provide feedback. Such assessment practices have a wide variety of learning benefits, including improved: content understanding (Reinholz, 2015b), communication (Reinholz, 2016), and self-assessment skills (Black, Harrison, & Lee, 2003; Sadler, 1989). While students learn from both giving and receiving feedback (Reinholz, 2015c), this paper focuses only on the *giving* aspect of feedback.

Not all feedback is equally useful (Hattie & Timperley, 2007; Shute, 2008). One productive lens for viewing feedback is that it helps a learner establish: (1) where they are, (2) where they going, and (3) what needs to be done to get there (cf. Black & Wiliam, 2009; Ramaprasad, 1983). From this perspective, it can be helpful for feedback to focus both on strengths and weaknesses, as this helps a learner establish their current status and possible goals (cf. Brown, 2012; Sadler, 1989). In addition, beyond simply identifying what is wrong, useful feedback helps a learner figure out how to improve. Given the close connection between feedback and learning, improving the quality of student feedback is likely to improve the quality of their learning, and thus is an important topic of focus.

The present study focuses on a particular peer assessment activity, called Peer-Assisted Reflection, or PAR (Reinholz, 2015b, 2015a, 2015c, 2016). PAR has been used in a variety of STEM disciplines (e.g., Reinholz, 2015b; Reinholz & Dounas-Frazer, 2016), but developed primarily in mathematics. PAR has four main components, used on a weekly basis: (1) students generate a draft solution to a homework problem, (2) students reflect on their drafts, (3) students exchange feedback with their peers, and (4) students revise their work before turning in their final submission. During the feedback exchange, students have five minutes to silently read each other's work and write comments, and then

https://doi.org/10.20429/ijsotl.2018.120210

five minutes for verbal conferencing. To support these conferences, students have a structured feedback form, and other various classroom activities associated with PAR (elaborated later). In what follows, I analyze three variations on the feedback forms and related activities to explore how they impacted the feedback provided by students.

METHOD

Context

Data for the present study were drawn from three iterations of PAR, two at the same institution (Reinholz, 2015b), and the third at a different institution (Reinholz, 2017). IRB approval was received for all of these studies. In each of the three iterations, PAR was used in an introductory college calculus course in which students completed a weekly PAR problem for homework. The students completed their peer conferences during class and then had until the next class session to revise their work and turn in a final submission. For reasons elaborated below, these three iterations are henceforth called the Questions (N = 56), Critique (N = 34), and Balanced (N = 124) approaches. As evident by the sample sizes, the Balanced approach was used in a large-lecture course (with four smaller corresponding recitation sections), compared to the Questions and Critique sections, which were smaller classrooms. While this had implications for pedagogy in general, it also meant that students in the Balanced section received no written feedback on their PAR assignments, whereas in the other sections the students received extensive written feedback on the quality of their solutions *and* their engagement with the PAR process (e.g., on their self-assessments or peer feedback). In this way, students in the Balanced approached did not receive an important support that the other students did.

Data Sources

Student solutions to PAR assignments were collected and scanned in all three approaches. Due to logistical constraints in the Balanced approach, student assignments could only be collected in two of the recitation sub-sections, which limited the overall amount of data collected. To support comparison between approaches, only common mathematical tasks were analyzed. Due to revisions of the tasks themselves, and differences in the content of calculus across the two institutions (i.e., one featured early transcendentals while the other did not), it was only appropriate to compare student solutions for three problems. This still provided a sizable 340 student solutions for analysis (see Table 1).

Table 1. *Number of solutions analyzed by problem and approach*

	Questions	Critique	Balanced	Total
Bottles	51	36	34	121
Ink Blot	51	25	33	109
Odd Function	48	29	33	110
Total	150	90	100	340

Design of Feedback Forms and Support

Following its theoretical goals, the purpose of PAR activities was to help students to provide supportive feedback to one another. An underlying assumption of PAR was that, especially in mathematics, it can be more difficult to identify weaknesses than strengths, so providing constructive criticism was emphasized. The idea was that as students learned to identify the weaknesses in their peers' solutions, they would be better able to identify their own weaknesses and improve upon them. To further support self-assessment, it was emphasized that students should learn to justify their answers, or have tools to see *why* their answers were correct, or not. Following a design-based approach, the PAR activities were revised across iterations, to better support these goals.

The Questions approach was the first iteration of PAR. The underlying rationale was that through regular exposure to a set of reflective questions, students would adopt these questions themselves and develop stronger self-assessment skills (cf. Schoenfeld, 1987). The instructor had a poster in the front of the room with three reflective questions and regularly used them during class. Students' PAR packets also included a page of reflective questions (on the back of the problem statement) to remind students of these questions. The page also had a set of self-assessment checkboxes that focused students on important aspects of their solution (e.g., explaining why, labeling graphs, use of pronouns). The peer feedback form itself had two main prompts:

- **Approach:** Give at least one suggestion to improve the communication/presentation of the solution. (You might focus on organization, explanation, labeling, etc. Be specific: don't say "it was hard to follow" or "the explanation was unclear" without saying *why* it was hard to follow, *what* was unclear, and *how* to improve it.)
- **Justification:** What evidence was provided that the problem was solved correctly? (Push your partner to justify "how they know;" also, note any errors that you found.)

There was also a third box for other optional feedback, which in practice students did not use often.

The Critique approach was developed in response to the study of student engagement with the Questions approach. In particular, interviews with students indicated that they rarely paid attention to the reflective poster in the front of the class and analysis of student work showed that students often simply said "everything looks good" even when they were repeatedly told to provide constructive criticism, not praise (e.g., there was not even a specific place on the feedback form for praise).

To address these issues, a new classroom activity that focused on critiquing the work of hypothetical students and having a whole-class discussion about the process was introduced (cf. Reinholz, 2015a, 2015b). The set of reflective questions was also removed from the PAR packets, and instead replaced with a set of hints for each problem. There were also minor changes in the wording on the self-assessment form and peer feedback form, which now read as follows:

- **Communication:** Give at least one suggestion to improve the communication of the solution. (Focus on explanations, imprecise use of language, organization, labeling, etc. Be specific: don't say "it was hard to follow" or "part 2 was unclear" without saying *why* it was hard to follow, *what* was unclear, and *how* to improve it.)
- **Correctness:** Note any errors you found. (Focus on misunderstanding of concepts, misuse of mathematical language, calculational errors, incomplete answers, etc. Be specific: don't just say "part 2 was wrong;" say exactly *what* is wrong, *why* it is wrong, and *how* to improve it.)

Once again, there was a third box for optional feedback.

The Balanced approach was the most radical set of revisions. Even though the prior feedback forms emphasized specific, critical feedback, students still provided a large number of general comments and praise. Thus, rather than pushing against students' desires to write positive comments, this approach embraced it. The new feedback form was divided into two halves: one for suggestions, and one for strengths. The rationale was that this would provide students with a place for identifying strengths in the solutions and also that by physically dividing the page in half, it would be clear that there was also a place for suggesting improvements.

In addition, rather than having separate self-assessment and peer feedback forms, the two were integrated into one. The middle of the PAR form featured a set of 10 icons related to key areas of focus (the same as in prior iterations). Rather than checking boxes, students were simply to circle icons that were associated with areas they wanted feedback on and cross out icons for areas that they did not want feedback on. The idea was that this would provide students with additional agency in choosing the categories of feedback that they wanted to receive, and still by repeatedly exposing students to these particular categories, it would help them use the categories to think about their own work. The feedback form is given in Appendix A. Like in the Critique case, students in the Balanced approach analyzed work as a whole class and discussed how to give feedback, but only for the first half of the semester.

Analytic Procedures

To prepare student work for coding, all work was de-identified and peer feedback forms were separated from the rest of students' PAR packets. To improve the accuracy of coding, student feedback was scored randomly across problems and approaches. While it was not possible to completely decouple the written work from which section it belonged to, mixing the work up would prevent systematic bias due to coder drift or related issues. (Because the forms looked different in the different approaches, and feedback sometimes included comments/graphs that could not easily be transcribed, it was not possible to completely separate coding from the original work.)

Each piece of feedback was coded along two dimensions: (1) the type of feedback and the (2) category of the feedback. Three types were chosen before coding began, aligned with the three key questions for learning from feedback (Black & Wiliam, 2009; cf. Ramaprasad, 1983): (A) identifying a strength, (B) identifying a weakness, or (C) suggesting a course of improvement. Feedback was coded as a strength when it recognized something positive about the solution or when a student indicated that they agreed with their peer. Feedback was coded as a weakness when it identified an area of disagreement or something that could be improved upon. If feedback identified a specific suggestion for how the solution could be made better, it was coded as an improvement. Because suggesting a course of improvement implicitly recognizes a weakness, if students did suggest improvements, feedback was only coded as category (C) and not a weakness (B). Thus, weaknesses were only coded in the instances when students did not also suggest how to improve.

Feedback was also coded along 11 different categories, which were aligned with the self-assessment and peer feedback forms used in PAR: (1) show all steps, (2) explanations, (3) pronouns, (4) mathematics vocabulary, (5) variables/units, (6) diagrams, (7) problem setup, (8) calculations, (9) multiple solutions, (10) answer, and (11) other feedback. Categories (1)-(6) focus explicitly on the communication of the solution, while (7)-(10) are more related to the correctness of the solution. These 10 categories were chosen as an *a priori* coding scheme, because they aligned with the types of feedback that students were encouraged to provide across all versions of the feedback form. The category "other" was added to capture other feedback that did not fall under these categories. For any given solution, a set of feedback could only be coded once along each category. Thus, if a student gave three pieces of feedback on how to improve a graph, it would be coded only once, under the category diagrams.

Examples of these feedback categories and types are given in Table 2. The examples are all taken from real student work, to show examples of things that students actually wrote. In some cases, students also provided graphical annotations (e.g., drawing a graph to suggest an improvement. Finally, it is noted that these particular categories align with the use of PAR in mathematics problem solving, but in other contexts, one could use other categories corresponding to their use of PAR. The purpose of looking at these particular categories was not to say that they are the ideal categories, but rather to see how the different forms influenced the spread of feedback given by students.

Table 2. *Examples of feedback categories and types*

Category	Example	Type
Show all steps	To make it easier to understand, write the problem itself next to each letter.	Improvement
Explanations	You could explain more.	Weakness
Pronouns	Don't use pronouns like "it." Say what you're referring to by name.	Improvement
Mathematics vocabulary	Used correct terminology	Strength
Variables/units	Be sure to include units.	Improvement
Diagrams	There is a part where the sides slope up diagonally, be sure to include that on the graph.	Improvement
Problem setup	You implied that f(-2) and f(2) were maximums, but the problem said that f(x) = 0 at x=-2 and x=2.	Improvement
Calculations	Calculations are good!!	Strength
Multiple solutions	Use the fundamental theorem of calculus to help you justify your answer.	Improvement
Answer	I don't think your answer is correct.	Weakness
Other feedback	Use more complex graphs to challenge yourself and prepare for the test.	Improvement

RESULTS

The results of analysis are separated into two subsections: quantitative results about the types and categories of feedback, and a brief qualitative description of the nature of feedback across approaches.

Types and categories of feedback

The distribution of types of feedback is given in Table 3. To make the results easier to interpret across sections with different sample sizes, the raw counts of feedback were divided by the total number of students in that approach. Thus, a value of 70% indicates that on average, 70% of students described a weakness in a peer solution. Similarly, the value of 145% indicates that, on average, students identified more than one strength in their feedback. In practice, some students may have identified multiple weaknesses, so more than 30% of students (in the Critique case) did not identify *any* weaknesses.

As the table shows, the feedback distributions for Questions and Critique were relatively similar, except that Critique students noted fewer strengths in their peers' work. In contrast, feedback provided through the Balanced approach featured *far* more strengths than the other two approaches. There was also a reduction in the number of weaknesses noted without suggestions for improvement, and the number of improvements was even slightly higher. These differences were significant, $\chi^2(4, N = 340) = 37.741$, $p = 1.27 * 10^{-7}$, with an effect size of $d = 0.71$.

Table 3. *Feedback types (as percentage of solutions)*

	Questions	Critique	Balanced
Strengths	73.3	56.7	145.0
Weaknesses	72.0	70.0	51.0
Improvements	79.3	87.8	92.0

Table 4 shows the distribution of feedback by category. For all three approaches, feedback focused primarily on explanations, diagrams, and answers. In the Questions and Critique approaches, the feedback was especially centered on explanations and the answer. This makes sense, because interpreted broadly, the two feedback boxes on the peer feedback form corresponded to communication and correctness (or explanations and the answer). While the Balanced approach also heavily emphasized these two areas, in general there was much more spread around the other categories. The increase in diagram-focused feedback for the Balanced approach can likely be attributed to the fact that this was an added category to the form not present in earlier iterations. These differences were significant (using Fisher's exact test to account for small values in the contingency tables), as $p < 10^{-7}$.

Table 4. *Feedback categories (as percentage of solutions)*

	Questions	Critique	Balanced
Steps	4.7	7.8	29.0
Explanations	81.3	77.8	77.0
Pronouns	2.7	1.1	6.0
Terminology	2.0	7.8	3.0
Units	1.3	3.3	9.0
Diagrams	28.0	31.1	63.0
Setup	2.7	2.2	18.0
Calculations	8.0	4.4	15.0
Multiple Solutions	2.7	2.2	5.0
Answer	86.0	75.6	46.0
Other	5.3	1.1	17.0

Figures 1, 2, and 3 show the distributions for strengths, weaknesses, and improvements, respectively. Figure 1 shows that the Balanced approach resulted in recognition of strengths across a wide variety of categories, whereas the other two approaches focused primarily on explanations and the answer. Moreover, across approaches, strengths focused on explanations and the answer tended be the most generic, often featuring statements like "good use of explanations," or "your answer looks solid." In contrast, when students articulated strengths in the other categories, it tended to be more specific.

Figure 2 shows the distribution of feedback focused on weaknesses. The distributions look relatively similar across approaches, except for the Balanced approach, which has far less feedback articulating weaknesses about the answer. This is consistent with the reduction in answer-focused feedback. In the other two approaches, feedback identifying weaknesses related to the answer generally indicated that the student providing feedback disagreed with the solution, or was not sure about the result.

Figure 3 shows the distribution of feedback focused on improvements. As before, the feedback in the Balanced approach tended to be more distributed across feedback types. Once again, the increase in diagrams feedback can likely be attributed to a change in the feedback form, explicitly drawing attention to this feature.

Quality of feedback

In addition to shifting the types and categories of feedback, the various approaches also resulted in feedback of a different quality. Feedback in both the Questions and Criticism approaches was generally narrative, consisting of complete sentences. In contrast, feedback from the Balanced approach was much shorter and pointed. Table 5 shows a comparison of feedback samples to highlight this contrast.

It is noteworthy that not all feedback in the Questions and Criticism approaches was written in narrative form (e.g., some students used bullet points). Still the presence of narrative statements in Questions and Criticism compared to the Balanced approach was striking. Even though some students in the Balanced approach did provide extended sentences of feedback, this was less common.

Table 5. *Comparison of feedback quality*

Questions / Criticism	Balanced
When drawing the graphs you may want to label graphs slightly better to spell out which graph is height and which graph is rate of change of height.	Maybe label the graphs a little more.
All the graphs need explanations because I don't understand why you drew the graphs a certain way.	Add an explanation to each graph.
One thing to add for tests would be labeling the x- and y-axis.	Label variables on graph.

Figure 1. *Distribution of feedback focused on strengths*

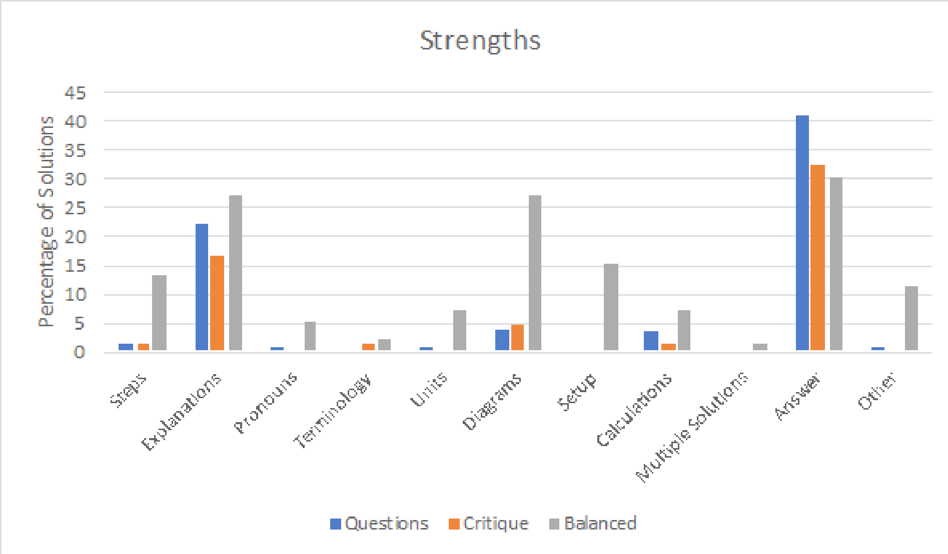


Figure 2. *Distribution of feedback focused on weaknesses*

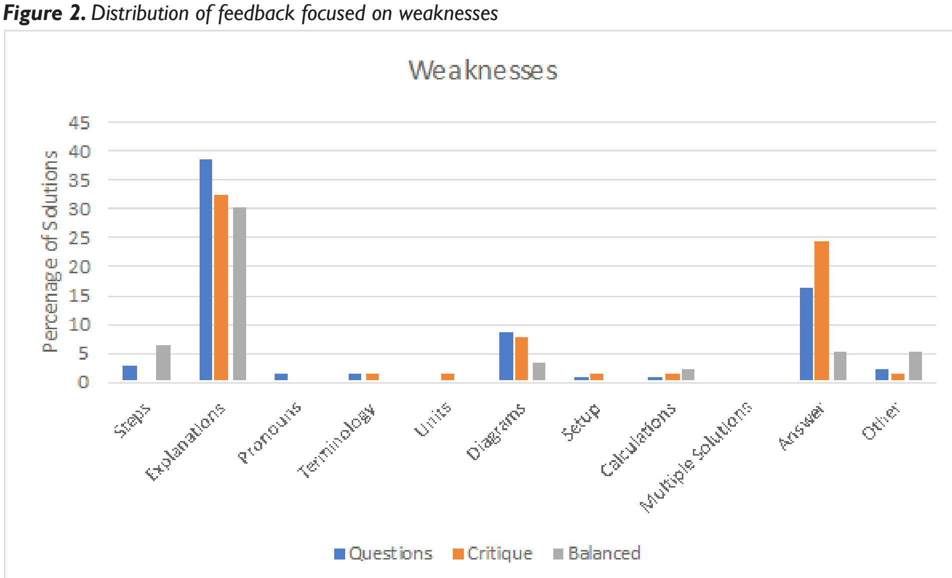
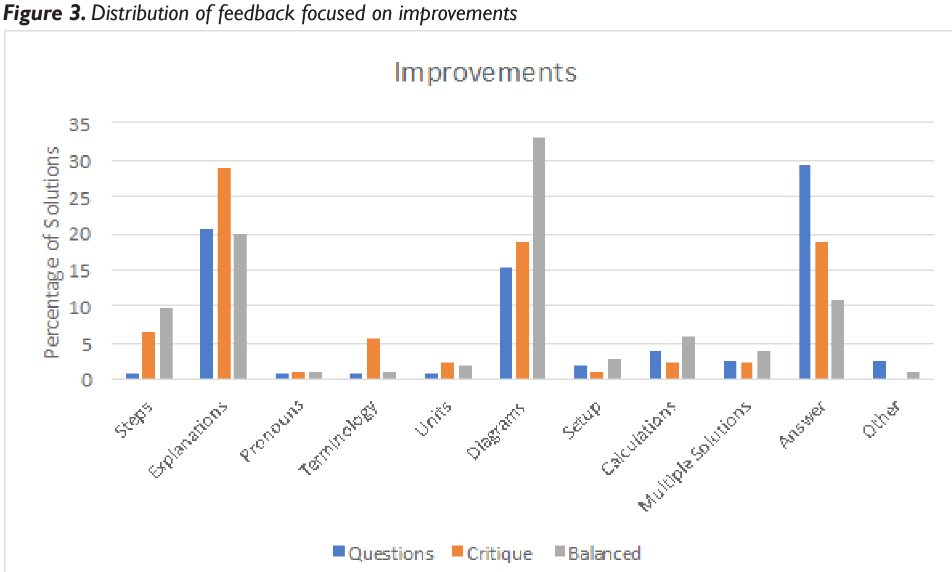


Figure 3. *Distribution of feedback focused on improvements*



DISCUSSION

Feedback is a key part of learning in general, and especially as a component of peer assessment. Yet, as this paper shows, the ways in which students are supported to provide feedback (e.g., through different artifacts) can have a profound impact on the types of feedback they provide. Thus, as educational designers, it is important for us to pay careful attention to how we organize such activities. In general, it is resource-intensive to develop and compare multiple, similar approaches. While this paper focused specifically on PAR, the results should generalize to other peer assessment contexts.

It is noteworthy that in every single approach, students articulated a number of strengths about their peers' solutions, even though they weren't explicitly prompted to in the Questions or Critique approaches. This appears to be a part of how humans interact: providing only critical feedback feels unnatural. Indeed, the colloquial term "compliment sandwich" has even been generated around this phenomenon: provide praise, criticism, and then praise, so that the positive ideas sandwich the negative one. The balanced approach embraced this tendency, providing space on the forms for students to articulate strengths, and also clearly separated them from the space for suggesting improvements. This was intended to remedy the potential problem of strengths "displacing" critical feedback in the feedback form, where students would feel as though they had completed the task without articulating improvements. As a result of emphasizing strengths, the Balanced students did articulate many more strengths than their peers, and they were more varied in their focus; moreover, there were less general statements such as "everything looks good."

Although the Balanced approach brought out more strengths-based feedback, it did not diminish students' articulation of courses for improvement. Although the differences were small, students in the Balanced approach actually articulated the largest number of improvements. In addition, the identification of weaknesses with no suggested courses for improvement was diminished. These types of comments are probably the least helpful, especially when given in generalities like "explain more," because they do not focus students on how to actually improve their explanations, and may just result in students writing excessively long but unclear explanations.

The Balanced approach spread out the feedback given by students to a number of different categories. There are likely multiple reasons for this. First, having a number of icons available on the feedback page makes the different categories much more salient than having them all embedded inside of longer sentences or on a previous page of a self-assessment form. Second, students who were receiving feedback chose the categories they wanted to receive feedback on, and these varied across all of the categories. In contrast, the default assumptions for the other forms were to focus on communication generally (which amounted to explanations) and correctness generally (which amounted to answers). This likely obscured the multi-faceted nature of communication (e.g., explanations, organization, labeling, diagrams) and correctness (e.g., setup, processes, answers.

There were also differences in the quality of feedback given. In the Questions and Critique approaches students generally wrote in narrative, giving complete sentences. Feedback in the Balanced approach was much more succinct. Because the form had a large number of categories on a single page, it may have unintentionally cued the students to write shorter feedback. In the context of PAR, this worked well because students had a chance discuss the feedback with one another. In some other contexts, the longer narratives may be preferable, as it could be easier to reconstruct meaning from them.

This study has its limitations. The contexts in which PAR was studied had differences in students, instructors, and the activities surrounding the artifacts. Moreover, without actually conducting in-depth user testing, one can only infer why students used the tools as they did. Finally, these studies took place in mathematics, but the results should generalize to other disciplines. Given that PAR has been used in a similar fashion in other STEM disciplines (e.g., physics, biology), the connection there is clear.

This study also opens up avenues for further research. Here the particular focus was on the feedback given by students. A future study could provide greater insight into how students actually interpret this feedback and how it supports their learning. Another key aspect of PAR is how it emphasizes both giving and receiving feedback. Both of these processes support learning, but likely in different ways. To the extent that future studies could disentangle giving and receiving feedback it would help further develop a theory of learning through peer assessment.

Ultimately, this paper cannot prescribe a single approach as the ideal way to design feedback forms. Instead, it provides a look into how students used a variety of artifacts and infers some principles about how artifacts may shape students' engagement. Each of these approaches has its virtues, and instructional design must always be related to a particular instructor's goals.

ACKNOWLEDGEMENTS

The author thanks Jason Siefken for creating the excellent graphics for the Balanced feedback form.

REFERENCES

Andrade, H. L., & Valtcheva, A. (2009). Promoting Learning and Achievement through Self-Assessment. *Theory Into Practice*, *48*(1), 12–19. https://doi.org/10.1080/00405840802577544

Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. Berkshire, England: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31.

Brown, B. (2012). *Daring greatly: How the courage to be vulnerable transforms the way we live, love, parent, and lead*. New York, NY: Gotham Books.

Cobb, P., Confrey, J., Disessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9–13. https://doi.org/10.3102/0013189X032001009

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, *24*(3), 331–350. https://doi.org/10.1080/03075079912331379935

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Herman, J., Epstein, S., Leon, S., Matrundola, D. L. T., Reber, S., & Choi, K. (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).

Min, H. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, *15*(2), 118–141.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: The National Council of Teachers of Mathematics.

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 133–148. https://doi.org/10.1080/0969594X.2013.877872

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, *28*(1), 4–13. https://doi.org/10.1002/bs.3830280103

Reinholz, D. L. (2015a). Peer conferences in calculus: The impact of systematic training. *Assessment & Evaluation in Higher Education*, 1–17. https://doi.org/10.1080/02602938.2015.1077197

Reinholz, D. L. (2015b). Peer-Assisted Reflection: A design-based intervention for improving success in calculus. *International Journal of Research in Undergraduate Mathematics Education*, *1*(2), 234–267. https://doi.org/10.1007/s40753-015-0005-y

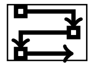
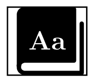
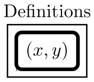
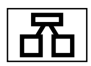
Reinholz, D. L. (2015c). The assessment cycle: A model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 1–15. https://doi.org/10.1080/02602938.2015.1008982

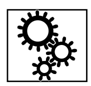
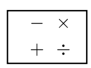
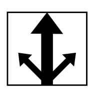
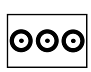
Reinholz, D. L. (2016). Improving calculus explanations through peer review. *The Journal of Mathematical Behavior*, *44*, 34–49.

https://doi.org/10.1016/j.jmathb.2016.10.001
Reinholz, D. L. (2017). Co-Calculus: Integrating the Academic and the Social. *International Journal of Research in Education and Science*, *3*(2), 521–542. https://doi.org/10.21890/ijres.327911
Reinholz, D. L., & Dounas-Frazer, D. R. (2016). Using Peer Feedback to Promote Reflection on Open-Ended Problems. *The Physics Teacher*, *54*(6), 364–368. https://doi.org/10.1119/1.4961181
Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714
Schoenfeld, A. H. (1987). What's all the fuss about metacognition? In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 189–215). Hillsdale, NJ: Lawrence Erlbaum Associates.
Schoenfeld, A. H. (2014). What Makes for Powerful Classrooms, and How Can We Support Teachers in Creating Them? A Story of Research and Practice, Productively Intertwined. *Educational Researcher*, *43*(8), 404–412. https://doi.org/10.3102/0013189X14554450
Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795
Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research*, *68*(3), 249–276. https://doi.org/10.3102/00346543068003249
Young, V., & Kim, D. (2010). Using assessments for instructional improvement.

APPENDIX A. BALANCED FEEDBACK FORM

Feedback Provided By: ____________________

Suggestions	Communication	**Strengths**
	 Show All Steps	
	 Explain Why, Not Just What	
	 Avoid Pronouns	
	 Use Correct Definitions	
	 Define Variables, Units, etc.	
	 Create Diagrams	

Suggestions	Accuracy	**Strengths**
	 Correct Setup	
	 Accurate Calculations	
	 Solve Multiple Ways	
	 Answer Reasonable	
	 Other (Write Below)	