International Journal of Engineering Education Vol. 34, No. 2(B), pp. 674–685, 2018
Printed in Great Britain

0949-149X/91 $3.00+0.00

Upping the Average: Manipulating Peer Feedback Quantity and its Effects on Feedback Quality*

GORDON G. KRAUSS
Department of Engineering, Harvey Mudd College, Claremont, CA, 91711, USA. E-mail: gkrauss@hmc.edu

ANDREA VASQUEZ, JOSEPH SINOPOLI and LAURA PALUCKI-BLAKE
Department of Engineering, Harvey Mudd College, Claremont, CA, 91711, USA.

Peer feedback during design reviews in engineering design courses can be limited in its effectiveness due to time limitations to solicit and respond to feedback and to self-censorship of comments. These actors tend to be prevalent in the traditional question and answer period following most in-class design reviews. The net effect is a reduced quantity of feedback on which the designers may take action. Prior work has shown a significant increase in the quantity of feedback offered in writing compared to that offered in a traditional question and answer (oral Q&A) format. The focus of this research is to investigate the ability of the instructors to manipulate the quantity of feedback offered by students during a design review by manipulating the expected number of comments during a design review. The quality of the feedback as a result of this manipulation is also evaluated by the design teams in terms of comment importance, professional tone of the comment, comment originality, and ease of use of the comment. The authors manipulated the expected number of comments upward from 2 to 4 across 3 sections of an introduction to design course over three design review exercises. The number of comments closely matched the number that students were told would be expected. The quality of the comments did not substantially differ across the sections, implying a greater quantity of comments can be obtained at negligible expense of quality by setting a higher expectation on the quantity of comments.

Keywords: peer review; feedback; written feedback; student comments; feedback quantity

1. Background

Interactive education requires the participation of students. Frequently, students in courses structured around project-based learning activities will be asked to engage in a feedback process for other students in the classroom setting with the goals of student reviewers learning to identify and communicate their observations and the students who are reviewed learning to receive and evaluate the feedback on their process and project. Providing and receiving feedback as a peer is a critical part of the design process. It has been observed that providing feedback can help facilitate student learning as it influences the quality of the peer reviewer's design artifact and presentation. Students are better able to apply feedback after having given feedback [1]. A greater quantity of peer reviewer comments provides a greater opportunity for improvement, or, in the case of repeated comments, provides reinforcement of the observed strengths or opportunities [2–4]. Many problem-based engineering design courses employ some version of a peer design review during student design exercises [5]. The goal of a classroom design review exercise for an engineering design course is to generate actionable feedback for the designers (presenters) as well as educate the reviewers on how to identify and share strengths and opportunities in the design process, resulting artifact and design review presentation [6–8].

Ideally, this peer feedback design review process would be objective. Ideal feedback is actionable; the ability of presenters to receive such feedback would be divorced from bias toward either the review process or resulting artifact. Feedback typically involves a presentation by the designers followed by a question and answer period by reviewers. The quantity of feedback offered by student reviewers in a design review of their classmates may be limited. This is due in part to the limited time available to provide feedback following a presentation but is also inhibited by the structure of many design reviews and the perceived social cost of offering comments. Often design reviews of student projects are scheduled such that they are temporally close meaning that a team that is being reviewed will soon transition to the role of reviewers. It may be that students are reticent to offer a candid review of a team in the presence of the instructor immediately prior to undergoing that review by their classmates [9]. It may be believed that critical comments would be perceived as hostile to the presenter's grades rather than helpful to development of their design skills and would be reciprocated during their own review [10]. Moreover, the social costs of perceived hostility will have ramifications outside of the classroom for students involved in the same course of study who are likely to study, live, socialize, and eat together and will be part of the same cohort even after graduation [11]. This is to say that in-class

* Accepted 12 January 2018.

criticism may have a cost that is believed by the reviewer to be much higher than any benefit to the design team. This limits the openness, quantity, and usefulness of the feedback the designers receive from their peers.

Additional factors that may affect peer feedback include gender of the reviewer, the student's learning style and anonymity of the review process [12–14]. Written feedback is demonstrated to mitigate some of the aforementioned effects and in doing so, increase the quality of feedback, making it a suitable alternative to traditional Q&A that suits the needs of this study [5]. It has been reported that online assessment is an effective way of acquiring written feedback, specifically learning-oriented peer feedback [15]. In online assessments, high quality prompts that guide the reviewer followed by the reviewer's high quality observations can increase the reflection levels of the presenter [16]. Online assessment can also effectively provide anonymity to reviewers.

Increasing quantity and quality of feedback may be presumed to be competitive as it could be assumed that a reviewer who can provide only a single piece of feedback would provide the most compelling comment or question while a reviewer providing two comments would add as the second comment the next most important comment. Thus, the average importance or impact of comments might be predicted to decline with increasing quantity. However, reviewers are not likely to mentally rank and categorize their feedback across the many ways each comment could be useful. Therefore, it is possible that many comments could be almost equally helpful or helpful in different ways. Establishing expectations regarding feedback quantity from each reviewer is one way that the amount of feedback may be increased. Understanding if this increases quantity and if that is at the expense of average quality is the focus of this study.

The quantity of feedback is important for getting a wide variety of feedback responses and ideas. This study investigates the impact of manipulating the number of comments a reviewer provides through establishing and sharing an average comment quantity per reviewer and by displaying a different number of feedback response boxes on the online forms used to provide comments in a design review.

Manipulation of student outcomes through establishing expectations has been investigated. Instructor's expectations can, in some cases, influence student growth. One investigation quantifies the degree to which teacher expectations influenced student growth was conducted by Rosenthal and Jacobsen [17]. Unarticulated teacher expectations may alter teacher investment in individual students or may raise the expectation of students themselves to perform at a higher level. In this study, clear expectations of satisfactory performance in terms of quantitative output were articulated to the student reviewers such that their performance with respect to the average could be measured. Left variable was the quality of the feedback as perceived by those students reviewed.

2. Study

Improvements with respect to quantity and quality of feedback have been observed when feedback is presented in written form rather than in oral question and answer sessions immediately following a design review presentation [5]. This was found to be true even when controlling for the time allocated to providing feedback. In this study, the ability of instructors to manipulate the quantity of feedback is investigated. It is expected that additional feedback from each reviewer will provide new perspectives on areas of a design process, resulting artifact, or presentation skills or will provide reinforcing feedback through repeated similar observations by different reviewers. Thus, the ability to increase the number of comments per reviewer through instructor manipulation will benefit the design team receiving feedback. In contrast, it might be expected that reviewers prioritize feedback such that the initial comments are most important and those presented later are less consequential, less original, less easily applied, or suffer from poor professional tone. That is, more comments mean a lower average quality of comment.

This study examines the ability of instructors to manipulate the quantity of feedback provided by student peer reviewers during a classroom based design review exercise. The variation in quality of feedback as a result of the difference in quantity is also evaluated.

The null hypotheses for this study are:

1. There will be no difference in the quantity (average number) of comments for each reviewer for each level of manipulation in the expected quantity of feedback.
2. The average quality of feedback as measured by the designers across the parameters of importance, uniqueness, ease of use, and professional tone will be statistically the same for all manipulation levels in quantity.

3. Method

3.1 Study setting and structure

This study was set at Harvey Mudd College, an all undergraduate liberal arts college. It takes place in an Engineering 004: Introduction to Engineering

Design and Manufacturing, an introductory, multi-disciplinary, undergraduate level design course composed of three separate sections, each with different professors but similar teaching structures. Each of the sections consisted of approximately 22 students. Every year, a preliminary design project (PDP) is assigned as a part of this project-based learning course and students in teams of three or four must follow a conceptual design process and present prototype progress in class. All sections held a design review lecture in which students presented their unfinished prototype to receive suggestions for improvement from the rest of the section using online written feedback forms. All sections also completed a final design review presentation to describe the process, artifact, and to characterize the performance of their project. A major design project (MDP) is completed by all students working on teams of 3 or 4 with different members than the PDP. The MDP projects differ between each team. Students presented a mid-point design review for the MDP in similar fashion to that competed for the PDP. Student feedback for all design reviews was solicited using an online form. For each design review, design teams had nine minutes to present their work and student reviewers had two minutes to enter their feedback using the online form.

Students provided feedback on the other five teams in their section and received and evaluated the feedback provided for their team. All students in each section completed tasks, presented, and received feedback in like manner. The only altered variable between sections was the number of comments students believed was the average number of actionable pieces of feedback per reviewer and the quantity of spaces immediately shown to them for comments in an online form. Each section was told a different number for the expected value and had a number of comment boxes shown on the screen equivalent to that number plus one. All students were given the option to add additional boxes in order to give additional comments (up to ten comments). Section 1 was given 3 boxes and told that the average number of comments per person per presenting team was 2, section 2 was given 4 boxes and was told that the average number of comments per person per presenting team was 3, and section 3 was given 5 boxes and was told that the average number of comments per person per presenting team was 4. All text boxes could be left empty. Each form asked a student to adopt a perspective that was to be used during the commenting process of a given team. The perspectives (positive, balanced, and critical) were distributed $\frac{2}{5}$ positive, $\frac{1}{5}$ balanced, and $\frac{2}{5}$ critical in all sections, so is not considered as a variable between sections and does not impact the results of this study. A sample prompt given to a team member for a presentation review in order to provide feedback is shown below.

> **Be sure that you are Section 2, Team 4** to use this form.
>
> Feedback during a design process is a critical tool to improve the outcome of the process, the design artifact, as well as the design process itself. The feedback process benefits the presenting team as well as the individual reviewers. Your feedback will be important to help the team improve both their design artifact and the design process they apply to this and future projects. Providing feedback will help you to develop the skills of identifying issues with a design process or artifact and communicating them to the presenting team.
>
> **Please enter one piece of feedback—Comment or Question—in a new entry box.** Please provide a sufficient quantity of actionable feedback for each team you review. *The average number of actionable, independent responses for a design presentation is 3 from each reviewer per presentation.*
>
> When providing feedback in a design review you will need to consider the information provided from different perspectives. One way of regarding the perspectives of a reviewer is the six "thinking hats" described by de Bono. These hats describe the way a presentation or review may be approached by those reviewing it. To develop skills across a set of thinking perspectives, you are asked to view presentations and to provide feedback with some of these different perspectives. Please pay attention to the perspective you are requested to consider for each review as it will change for the presenting teams.
>
> Your feedback will be evaluated with respect to quantity and requested perspective. Please be sure to understand the perspective for each team you are reviewing and to enter each comment or question in a new field.
>
> *Comments on Team 1 presentation*
>
> Please adopt a [positive/balanced/critical] feedback perspective for this review. As a reviewer applying this perspective, you will try to identify [strengths/both strengths and flaws or shortcomings/flaws or shortcomings] in the design artifact or design process of the presenting team. The team is likely to have missed these [areas of strength/areas/errors or omissions] and the [positive/balanced/critical] feedback perspective review will indicate what [should be continued to maintain /should be continued and must be done better to continue to improve/must be done better to continue to improve] the outcome of this specific design and to improve the design process

in the future. You should include all observations in your comments, but should try to capture all [strengths/strengths and flaws/flaws] in the process.

Prior to completing the form, students were asked what the average number of actionable responses for a design presentation such as the one they are reviewing is to check for comprehension of the instructions. For each comment, student reviewers were asked to identify the primary topic of the comment (presentation, design process, design artifact, or other) and to rate their comment on a five-star scale of importance judged at their own discretion. All sections were given the same amount of time to present their work (9 minutes) and sections were grouped by course section (randomly assigned by the registrar). Data from this study comes from the feedback from the PDP design review, the PDP final presentation, and the MDP design review.

Design teams received their feedback within a day of presenting. Each student on a design team evaluated every piece of feedback with respect to comment importance, comment originality, ease of use of the comment, and comment professional tone. Comments were evaluated on a 5–point Likert scale (1—Strongly Disagree, 2—Disagree, 3—Neutral, 4—Agree, 5—Strongly Agree).

The study controlled factors that might lead to improved feedback from students while attempting to hold constant the class time used for the design presentation. Thus, differences in feedback are not attributable to (1) The total class time spent on presentations and feedback, (2) The degree to which the instructors encouraged feedback and the specificity of feedback requested, (3) The means by which students recorded and submitted their feedback, (4) The means by which the student teams presented their design reviews, and (5) The method of interaction between teams and reviewers to discuss feedback. Additionally, oral comments and questions were not permitted from faculty or students following the design reviews to prevent reviewers becoming biased toward a comment from an oral reviewer.

3.2 Metrics used

Feedback was measured in multiple ways. Important metrics used to investigate the effect of included the quantity of comments, the ease of use of the comments as perceived by the presenter, the professional tone of the comments as perceived by the presenter, the originality of the comments as perceived by the presenter, and the importance of the comments as perceived by the presenter. The quantity of comments is determined through a summation of all comments in a given section. The number of comments per reviewer divides the total number of comments by the total number of reviewers. Quality of comments in each category are evaluated using 5 point Likert scale with the following prompt:

Feedback Evaluation
Some feedback is more useful than other feedback. To help you receive feedback you will need to individually and without consulting with your team rate each comment on four scales: ease of use, tone, originality, and importance. After this you will need to identify (but not rank) the three best and worst comments. For purposes of this evaluation, please apply the following criteria for evaluation:

1. Ease of use: Is this comment easy to apply to the design you are generating or to improving your future design process?
2. Tone: How well does this comment convey a professional tone?
3. Originality: How frequently do you see similar comments to this one?
4. Importance: How important is applying this comment to the success of your project?

4. Results

4.1 Comment quantity

Student reviewers were manipulated in terms of the number of pieces of average feedback expected during a written design review. Each section was told that a different number of comments was the average number for a student reviewer to submit. Specifically, section 1 was told that 2 comments were average, section 2 was told that 3 comments were the average, and section 3 was told that 4 comments were average. In addition, each section had an online submission form that showed one more text entry box that nth average number of comments reported to them. All sections were able to enter more comments by selecting an "add more comments" button at the end of the last text entry box. A graph showing the quantity of feedback collected for the design review of Preliminary Design Project, the final presentation of the Preliminary Design Project, and the design review of the Major Design Project is given in Fig. 1.

The average number of comments per student reviewer was 2.24 in Section 1, 3.19 in Section 2, and 4.0 in Section 3 during the PDP first design review. The average number of comments per student reviewer was 2.34 in Section 1, 3.33 in Section 2, and 4.4 in Section 3 during the PDP final design review. The average number of comments per student reviewer was 2.22 in Section 1, 3.17 in Section 2, and 4.09 in Section 3 during the MDP

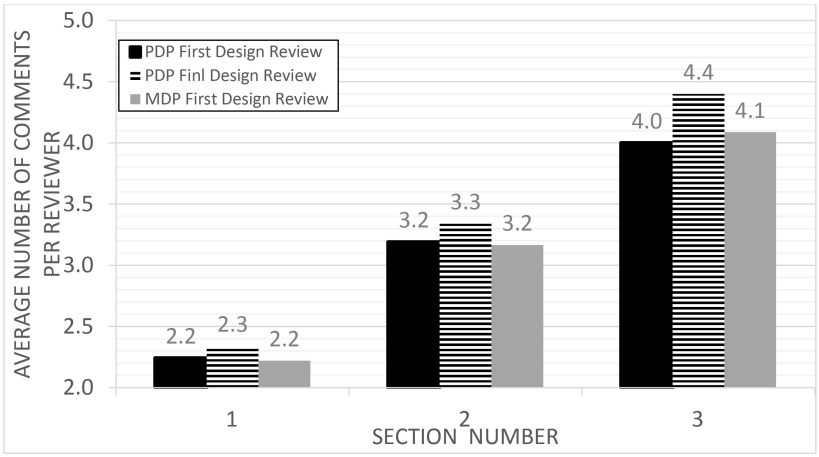

Fig. 1. Feedback Quantity by Section. Average number of comments per reviewer for each section. As part of the study manipulation, each section was told a different number of actionable comments per reviewer was typical: Section 1 was told 3 comments, Section 2 was told 3 comments, and Section 3 was told 4 comments were typical. In fact, the average number was closer to 3 (average of 3.42 comments for written feedback in prior study).

first design review. The difference in quantity of feedback is statistically significant between sections for each trial of the experiment ($p < 0.0001$). It is noteworthy that within each test, the data fits a linear trend with a high degree of fit ($R^2 > 0.998$ on all cases).

4.2 Designer rating of comment quality

There are many ways to evaluate the quality of comments received during feedback. This study required each member of a design team receiving feedback to evaluate each comment received across four topics. Each design student rated each piece of feedback with respect to ease of use, professional tone, importance, and originality.

4.2.1 Ease of use of comments

To characterize the ease of use of comments, each design student was prompted with this question for each comment received: Is this comment easy to apply to the design you are generating or to improving your future design process? Answers were recorded on a 5-point scale: "This comment was [not/ somewhat/ moderately/ significantly/ exceptionally] easy to use." Fig. 2 shows the resulting ease of use of comments across all sections for each of the three trials of the study.

It can be seen in Fig. 2 how the ease-of-use scores vary between sections for the manipulation of the expected number of pieces of feedback. For the Preliminary Design Project Design Review, the

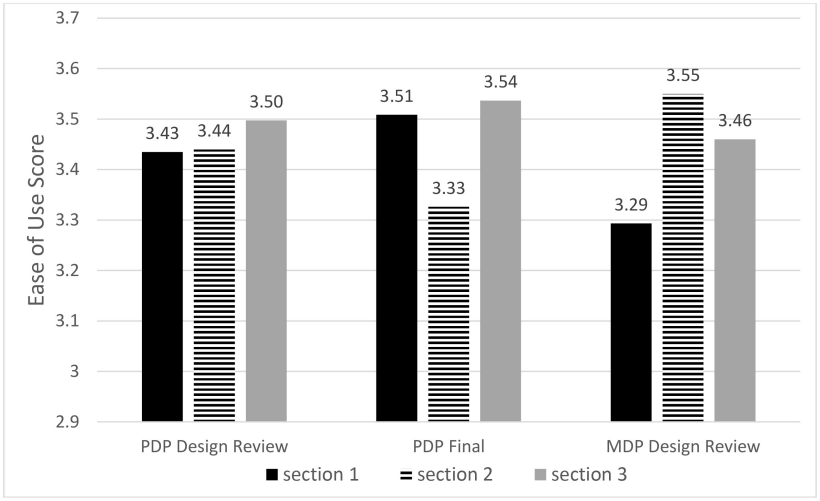

Fig. 2. Average Ease of Use of Comments Received. Each design student rated each comment they received on its ease of use. The resulting average scores by section are shown.

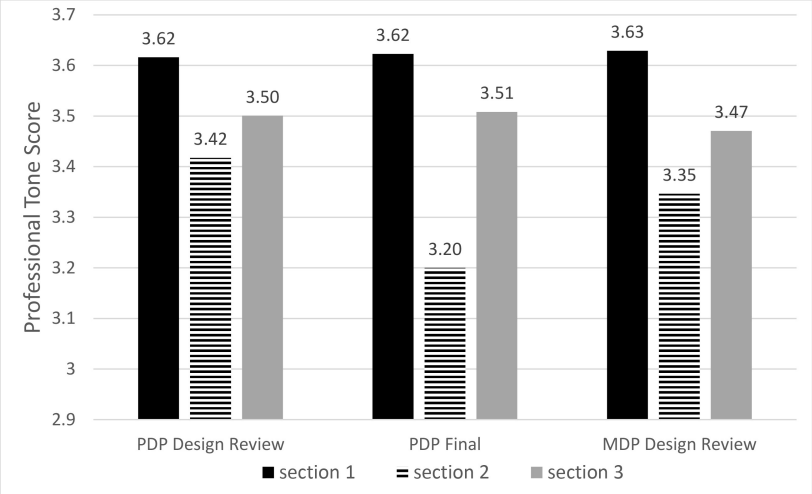

Fig. 3. Average Professional Tone of Comments Received. Each design student rated each comment they received on its professional tone. The resulting average scores by section are shown.

differences between sections are not statistically significant (section 1–2, $p = 0.95$, section 2–3, $p = 0.41$, section 1–3, $p = 0.44$). For the Preliminary Design Project Final Presentation, the differences between sections are statistically significant between sections 1 and 2 ($p = 0.0234$) and sections 2 and 3 ($p = 0.0065$), but not between sections 1 and 3 ($p = 0.7126$). For the Major Design Project Design Review, the differences between sections are statistically significant between sections 1 and 2 ($p = 0.0023$) and sections 1 and 3 ($p = 0.0565$), but not between sections 2 and 3 ($p = 0.2469$).

4.2.2 Professional tone of comments

To characterize the professional tone of comments, each design student was prompted with this question for each comment received: How well does this comment convey a professional tone? Answers were collected on a 5-point scale: "This tone is [not/ somewhat/ moderately/ significantly/ exceptionally] professional." Fig. 3 shows the average professional tone score of comments across all sections for each of the three trials of the study.

It can be seen in Fig. 3 how the professional tone scores vary between sections for the manipulation of the expected number of pieces of feedback. For the Preliminary Design Project Design Review, the differences between sections 1 and 2 are statistically significant ($p = 0.0191$) but are not statistically significant between sections 2 and 3 ($p = 0.3690$) or between sections 1 and 3 ($p = 0.2009$). For the Preliminary Design Project Final Presentation, the differences between sections are statistically significant between sections 1 and 2 ($p < 0.0001$) and sections 2 and 3 ($p = 0.0007$), but not between sections 1 and 3 ($p = 0.1738$). For the Major Design Project Design Review, the differences between sections are statistically significant between sections 1 and 2 ($p = 0.0012$) and sections 1 and 3 ($p = 0.0527$), and between sections 2 and 3 ($p = 0.0526$).

4.2.3 Originality of comments

To characterize the originality of comments, each design student was prompted with this question for each comment received: How frequently do you see similar comments to this one? Answers were recorded on a 5-point scale: "This tone is [very common/ common/ is neither common nor uncommon/ is uncommon/ is unique]." Fig. 4 shows the average originality score of comments across all sections for each of the three trials of the study.

It can be seen in Fig. 4 how the originality scores vary between sections for the manipulation of the expected number of pieces of feedback. For the Preliminary Design Project Design Review, the differences are statistically significant between sections 2 and 3 ($p = 0.0563$) but not between sections 1 and 2 ($p = 0.3324$) or between sections 1 and 3 ($p = 0.5635$). For the Preliminary Design Project Final Presentation, the differences are statistically significant between sections 1 and 2 ($p = 0.0042$) and between sections 2 and 3 ($p = 0.0185$) but are not statistically significant between sections 1 and 3 ($p = 0.2698$). For the Major Design Project Design Review, the differences between sections are statistically significant between sections 1 and 3 ($p = 0.0160$) but not between sections 1 and 2 ($p = 0.2441$) or between sections 2 and 3 ($p = 0.1200$).

4.2.4 Importance of comments

To characterize the importance of comments, each

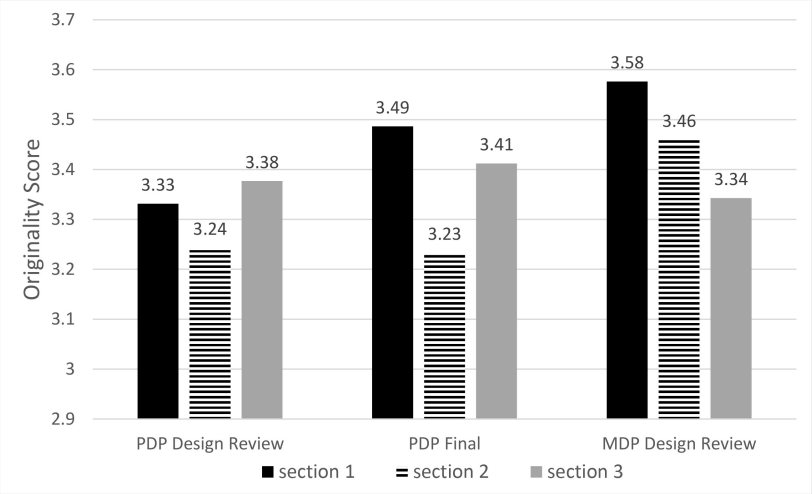

Fig. 4. Average Originality of Comments Received. Each design student rated each comment they received on its originality. The resulting average scores by section are shown.

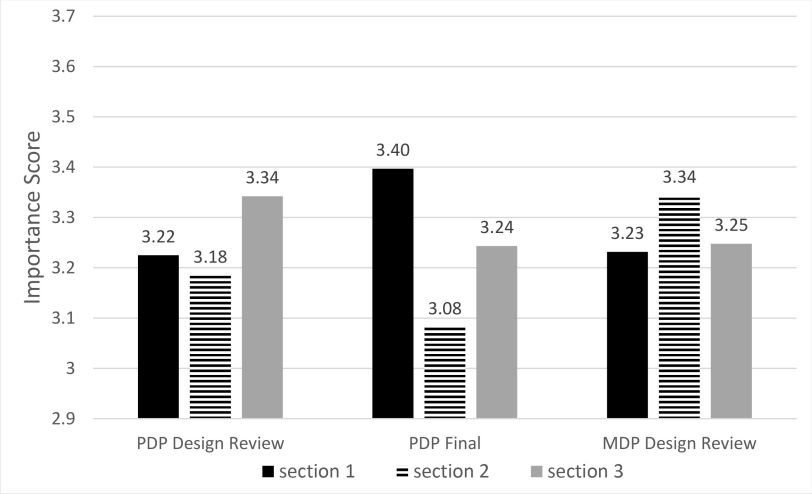

Fig. 5. Average Importance of Comments Received. Each design student rated each comment they received on its importance. The resulting average scores by section are shown.

design student was prompted with this question for each comment received: How important is applying this comment to the success of your project? Answers were recorded on a 5-point scale: This comment is [not/ somewhat/ moderately/ significantly/ exceptionally] important."

It can be seen in Fig. 5 how the importance scores vary between sections for the manipulation of the expected number of pieces of feedback. For the Preliminary Design Project Design Review, the differences are statistically significant between sections 2 and 3 ($p = 0.0494$) but not between sections 1 and 2 ($p = 0.6045$) or between sections 1 and 3 ($p = 0.1417$). For the Preliminary Design Project Final Presentation, the differences are statistically significant between all sections: 1 and 2 ($p = 0.0001$), 2 and 3 ($p = 0.0258$), and 1 and 3 ($p = 0.0263$). For the Major Design Project Design Review, the differences between sections are not statistically significant between sections 1 and 2 ($p = 0.2317$), between sections 2 and 3 ($p = 0.1487$), or between sections 2 and 3 ($p = 0.8564$).

5. Discussion

5.1 Comment quantity

The manipulation applied to the quantity of expected comments applied in this study consisted of two parts. First, an expected number of comments from each reviewer was established. Second, the number of text boxes immediately available for reviewers was equal to one more than the number of expected comment. Section 1 was told the average number of comments per reviewer is two and was provided three text boxes on their initial review page. Section 2 was told the average number of comments per reviewer is three and provided four text boxes on their initial review page. Section 3 was

told the average number of responses per reviewer is four and was provided five text boxes on their initial review page. Every section was given the option to add more comments through a question at the end of the page. All text boxes were optional, meaning no feedback was required to complete the form. In similar prior studies, the expected number of written comments per review was 3.42.

Manipulation of the number of comments was statistically significant between all sections and was repeatable with different design reviews and for different projects as shown in Fig. 1. This manipulation is effective in increasing or decreasing the number of comments compared to those provided in an un-manipulated prior study. Because time to respond was controlled across all sections, the number of comments does not reflect an increase in the time available to reviewers and no additional motivation or encouragement was given for the different sections.

In the event where the number of comments is decreased compared to the typical expected (as for sections 1 and 2), the quantity closely tracks the stated value. That is, on average each reviewer provided slightly more—about 0.2 comments more—than the expected number of comments. Given that each reviewer gives feedback on five teams in a section, 0.2 more comments per review equates to 1 extra comment per reviewer per section, on average. For sections 1 and 2, the number of comments submitted per reviewer per team reviewed was lower than the normal number that would be expected without the manipulation.

For the case where a higher number of comments were set as the expectation, section 3, the average number of comments were on average 0.16 comments greater than the expected number, but this was driven by a higher average number of comments (4.4) in the PDP final presentation combined with lower response rates in the PDP design review (4.0) and MDP design review (4.1). It is unclear if the number of comments is a result of the type of review or if this is a normal variation in the response rate. Setting a higher expectation than the previously observed value resulted in more feedback per reviewer.

The number of comments made by each reviewer tracks well with established expectations over the range studied. In these studies, there was an impressive linearity between the sections for each of the different test series. Instructors may apply this finding in their work by setting a higher expectation for the number of design review comments expected from student reviewers. The observed linearity cannot extend indefinitely as the number of comments as the number of comments is finite as is the time available to generate the comments. Nevertheless, instructors should be aware that their articulated expectations have significant impact on the number of comments submitted.

5.2 Comment quality

Though additional comments can be generated as a result of setting higher expectations, the quality of the additional comments is an important consideration. If the added comments are of lower quality, there will be a diminishing return on the increased expectations. If additional comments are merely 'filler' to complete the assignment, the value will be low. But if the additional comments are of equivalent quality, then it is reasonable to conclude that the expectation of higher quantity results in not just more comments but also in more comments that are useful.

5.2.1 Quality of comments for ease of use

It might be expected that the ease of use of the comments submitted by reviewers decreases as more comments are submitted. That is, one could assume that reviewers might first submit the most easily actionable comments and only submit the less easily applicable comments when those easiest to achieve have been submitted. Examination of Fig. 2 does not show a significant trend in terms of the ease of use for comments submitted by section. That is, establishing a higher expectation for the number of comments does not result in a reduced ease of use of the comments.

With respect to comment ease of use, there appear to be comments of equivalent quality unarticulated by reviewers that can be extracted through setting a higher expectation in quantity. While the observed ease of use consistency is true for the range of comment quantity examined in this study, it cannot be known if this behavior would continue indefinitely. It may be that harder to apply comments are more frequently submitted with an increased quantity of submitted comments than those studied in this experiment. Alternatively, easy to apply comments may increase with a higher number of submitted comments than those studied.

5.2.2 Quality of comments for professional tone

Establishing a higher expectation on the number of comments from student reviewers might be expected to result in a sacrifice in terms of the professional tone of the comments. This could be expected as a result of a rush to communicate the increased volume of information creating an atmosphere where brusque comments are submitted, or the tone is ill-considered. Examination of Fig. 3 indicates that there may be one case—the Major Design Project—where an increasing number of

comments correlates with a decreasing professional tone. However, this trend is observed between the lowest number of comments (section 1) and the highest number of comments (section 3) with a difference of 0.16 on a five point Likert scale. Though directionally lower in the first two studies with similar values (PDP design review at 0.12 and PDP final presentation at 0.11), these differences were not statistically significant. Moreover, section 2 tends to underperform in all sections in terms of professional tone, resulting in a lowered score for the middle number of expected comments. The results indicate that a decrease of professional tone is not predictable based on the expected feedback quantity.

5.2.3 Quality of comments for originality

It may reasonably be expected that an increasing quantity of comments will experience an increasing overlap in content, thus, a reduction in comment originality. Examination of Fig. 4 may demonstrate a downward trend in comment originality for one of the three tests conducted (the MDP decreases by 0.24 from section 1 to section 3). The other two sections do not show this trend statistically between the lowest (section 1) and highest (section 3) number of comments. As observed in Professional tone of comments, section 2 tends to average below the other sections in the PDP design review and PDP final presentation. It is interesting that the average comment originality remains approximately the same despite approximately doubling the number of comments.

5.2.4 Quality of comments for importance

The expectation that reviewers will provide the most important comments first is not supported by the data in Fig. 5. While the comment importance varies between sections, the only test case with a statistically significant drop between the first and third sections is the PDP final presentation with a decrease of 0.16. There is no statistical difference between sections one and three for the PDP design review or the MDP design review. While it is possible that a decrease in importance is observed with increasing comment quantity, the decrease is marginal compared to the additional number of comments.

5.2.5 Quality of comments section differences

It is worth noting that there is an apparent difference between section 2 and the other two sections for some of the quality questions (in particular, for comment tone, and, to a lesser degree, comment importance). The scores for section 2 tend to be lower than the other sections while it might reasonably be expected that they would fall between the other extremes. This may be due to the natural variation between section performance though it may be due to an uncontrolled factor that is unclear to the authors.

5.3 Issues of generalizability and thoughts for educators

5.3.1 Generalizability

The sample size of teams in each section of this study (6) and the number of students in each section of this study (n of approximately 22) are not unusual for design courses, though team and section sizes can vary with institution. Differences between test conditions were examined for statistical difference through the use of student t-tests (assuming heteroscedastic, two-tailed behavior). The differences noted as statistically significant are meaningful due to the amplitude of the difference between the mean values and include consideration of the standard deviations of the populations. Though larger sample sizes are desirable to increase sensitivity to differences, such populations may not be possible within the context of this course as it is strongly desired to maintain a low student to faculty ratio. Similarly, this study was conducted in an introductory design course and the differences so noted may or may not persist in upper division courses or in other settings of introductory design courses. Nevertheless, it is noteworthy that the differences are statistically significant and that methods to increase the sample size, possibly through use of combining section scores, in future work are of interest to the authors.

5.3.2 Thoughts for educators

Given the issues presented with respect to generalizability of these findings, educators may wish to adapt part, or all of the techniques presented here to increase fluidity and quality of feedback during peer reviews in design courses. It is possible, but unstudied, if such techniques are applicable beyond design courses with similar outcome. Application of the tools used in this study, specifically on-line survey systems, to capture and analyze the feedback of reviewers are costly in both time and effort and may not be necessary to achieve some of the desired outcomes. The tools used enabled a deeper study of the feedback from each reviewer to understand contributions that might otherwise have gone unattributed. There are a number of free online survey tools that could be modified to solicit feedback in the method suggested by this study. Though development of these tools for a given course would require substantial effort, this remains an option. It is the hope of the authors that an easy to use version of the tools used in this study will eventually be made available to the education community.

Nevertheless, index cards (either anonymous or signed) provided to reviewers will likely increase the number of comments obtained and manipulation specifically citing a number of expected comments might be used to encourage a higher number than would otherwise be delivered. It is also possible that other methods may prove effective in soliciting greater a greater quantity of feedback. The authors believe it likely that written feedback delivered in written rather than online form will exceed the quantity and quality of traditional oral question and answer feedback.

The authors recognize a risk with anonymized written feedback is that comments may be included that could be considered inappropriate or objectionable in phrasing or content. It appears to be the case that reviewers are comfortable writing anonymized comments that they would never say aloud. This might be addressed in multiple ways, particularly in the event anonymity has been assured to the reviewers. Prior to soliciting feedback, the educator may wish to caution reviewers on phrasing and appropriateness of comments and may want to spend time coaching how to give effective feedback. In the event comments are received that require a response, the instructor may decide to maintain anonymity and address ways the comment could have been better phrased as a class exercise. Of course, threatening, harassing, or other dangerous comments should be addressed by the appropriate parties.

5.4 Future work

It is evident from this work that the number of comments provided by student reviewers in a design course can be influenced significantly over the range examined without compromising overall quality of comments to a great degree. The limit to this range of influence without compromising comment quality should be further explored as well as the factors which influence it. The expected quantity of comments is clearly a driving factor in such a study. Additional factors may include the time allotted to enter comments, the motivation from instructors for providing feedback including specificity of topic, and the ability of the course to track and evaluate feedback by each reviewer.

There was no training provided to student reviewers in this study with respect to providing comments that were written in a professional tone, to generating or reporting comments that were original, to identifying easy to use comments, and to identifying important comments. Establishing interventions to improve the ability of reviewers to provide better feedback would be of interest for design reviews. Similarly, interventions aimed at improving the ability of designers to receive, evaluate, and apply feedback would be of benefit to design students. For example, it may be possible to train students to separate message tone and content to have a more objective evaluation and application of comments that are not phrased well but contain helpful information.

As a result of the process applied here, the number of comments has increased dramatically compared to the more standard oral question and answer period following a design review. While the increase in comments is welcome in terms of their utility to designers, the increase in the number of comments can make it difficult to process all the content. Additional information that might provide an easier examination of the increased quantity of comments by designers includes an ability to rank the comments by importance, topic, or originality prior to review. It would be of interest to develop a process able to direct designers to the appropriate comments in order of applicability to their needs. As a first step, a simple sorting of comments as evaluated by the reviewers themselves may help designers to prioritize their actions.

Separating the effect of the number of boxes and the stated comment average rather than in combination to determine whether the number of boxes, the stated average, or a mixture of the two is really driving the result is of interest to the authors. It would also be worthwhile to look into other variables that could have potentially caused the decreases in section 2. In a separate project, it might also be useful to look into what personal and individual factors influence comment, ease of use, importance, originality, and tone and what other factors might significantly influence the perceived comment quality.

Additional factors such as the impact of social relationships on the feedback process, the role of gender, the perception of comments due to topic or perspective, and the perceived degree of anonymity are worthy of investigation. While these factors are less easily manipulated than some of those previously indicated, they could conceivably play a larger role in the candidness of the feedback.

6. Conclusion

It is possible in design courses to alter the number of student review comments through manipulation of the expected number of comments and the initial number of text boxes available. This manipulation can increase or decrease the number of comments provided by the student reviewers to the presenting design teams. This finding suggests that instructors should carefully consider the expectations set for their design reviews in terms of the expected quantity of comments. The instructor expectation com-

bined with the written medium of review lends itself to considerable influence over the number of comments a design team will receive.

While statistically significant changes between sections with different comment quantity were sometimes observed, and while these changes sometimes resulted in lower quality of comments on average, the decrease was neither repeatable across the tests for the three presentations observed (from PDP design review to PDP final presentation to MDP design review) nor did the data for the center case continue the trends in all cases. The authors believe the differences to be small (on the order of 0.16 points on a Likert scale) compared to the increase in comment quantity (an average increase of 1.93 comments per reviewer) almost doubling the number of comments. As a result, the authors conclude no significant change in quality of comments as evaluated for professional tone, originality, importance, or ease of use was observed for increasing or decreasing the expected number of comments. That is, there is no increase in comment quality resulting from a lower expectation in comment quantity and there is no decrease in comment quality resulting from a higher expectation in comment quantity.

This study was conducted using an online form to provide feedback which was shared with the presenting design team after the presentation concluded. It may be that the benefits of written feedback persist with written forms. Additionally, the benefit of this feedback format may extend to course outside of engineering design or engineering where feedback from students is encouraged or required. The increase in feedback quantity achieved by enabling (and expecting) feedback from each reviewer in writing rather than those self-electing to share their thoughts should be possible in any course requesting feedback. It is possible that the manipulation in terms of quantity and the consistency in terms of quality may also translate.

References

1. L. li, X. Liu and A. L. Steckelberg, Assessor or Assesse: how student learning improves by giving and receiving feedback, *British Journal of Educational Technology*, **41**(3), 2010, pp. 525–536.
2. K. Topping, Peer assessment between students in colleges and universities, *Review of Educational Research*, **68**(3), 1998, pp. 249–276.
3. I. van den Berg, W. Admiraal and A. Pilot, Designing student peer assessment in higher education: analysis of written and oral peer feedback, *Teaching in Higher Education*, **11**(2), 2006, pp. 135–147.
4. Y. Xiao and R. Lucking, The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment, *Internet and Higher Education*, **11**, 2008, 186–193.
5. G. G. Krauss and L. Neeley, Peer Review Feedback in an Introductory Design Course: Increasing Student Comments and Questions through the use of Written Feedback, *International Journal of Engineering Education*, **32**(3B), 2016, pp. 1445–1457.
6. S-C. Tseng and C-C. Tsai, On-line assessment and the role of peer feedback: A study of high school computer course, *Computers and Education*, **49**, 2007, pp. 1161–1174.
7. Y-T. Sung, K-E. Chang, S-K. Chiou and H-T. Hou, The design and application of a web-based self- and peer-assessment system, *Computers and Education*, **45**, 2005, pp. 187–202.
8. S. Gielen, E. Peeters, F. Dochy, P. Onghena and K. Struyven, Improving effectiveness of peer feedback for learning, *Learning and Instruction*, **20**, 2010, pp. 304–315.
9. Y. Kali, and M. Ronen, *Design principles for online peer-evaluation fostering objectivity*, CSCL '05 Proceedings of the 2005 conference on Computer support for collaborative learning: the next 10 years!, 2005, pp. 247–251.
10. D. Magin, Reciprocity as a Source of Bias in Multiple Peer Assessment of Group Work, *Studies in Higher Education*, **26** (1), 2001, pp. 53–63.
11. N. A. Gennip, M. S. Segers and H. H. Tillema, Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions, *Learning and Instruction*, **20**(4), 2010, pp. 270–279.
12. A. M. Vasquez, S. Silcox, J. Sinopoli, L. Palucki-Blake and G. G. Krauss, Modes of feedback in design review process: Implications for utility and effectiveness based on student gender, *American Society for Engineering Education Proceedings*, Tempe, Arizona, April 20, pp. 2–17.
13. S. Lin, E. Liu and S. Yuan, Web-based peer assessment: feedback for students with various thinking styles, *Journal of Computer Assisted Learning*, **17**, 2001, pp. 420–432.
14. A. J. Bloom and J. E. Hautaluoma, Effects of message valence, communicator credibility, and source anonymity on reactions to peer feedback, *The Journal of Social Psychology*, **127**(4), 2001, pp. 329–338.
15. M. Keppell, E. Au, A. Ma and C. Chan, Peer learning and learning-oriented assessment in technology-enhanced environments, *Assessment & Evaluation in Higher Education*, **31**(4), 2006, pp. 453–464.
16. N-S. Chen, C-W. Wei, K-T. Wu and L. Uden, Effects of high level prompts and peer assessment on online learners' reflection levels, *Computers & Education*, **52**, 2009, pp. 283–291.
17. R. Rosenthal and L. Jacobsen, *Pygmalion in the classroom: teacher expectation and pupils' intellectual development,* New York: Holt, Rinehart and Winston, 1968

Gordon G. Krauss is the Fletcher Jones Professor of Engineering Design in the Department of Engineering at Harvey Mudd College. His design research interests include improving the way designers interact with each other in the design process and how design process tools are applied. Prior to joining Harvey Mudd College, Dr. Krauss was a lecturer in Mechanical Engineering at the University of Michigan and enjoyed a career in industry. He holds a PhD in Mechanical Engineering, an MS in Aerospace Engineering, both from Boston University, and completed his undergraduate degree in Physics and Astronomy at Haverford College.

Laura Palucki Blake is the Director of Institutional Research and Effectiveness at Harvey Mudd College. Her primary role is to coordinate data collection, interpretation and dissemination to support teaching and learning, planning and decision-making across the college. Prior to joining Harvey Mudd, she was the Assistant Director of the Cooperative Institutional Research Program (CIRP) at the Higher Education Research Institute (HERI) at UCLA. At UCLA, her primary

responsibilities included working with institutions to both implement various CIRP surveys and to develop and communicate about effective and innovative ways to using data to promote institutional improvement. Laura is a Teagle scholar at Wabash College's Center of Inquiry into the Liberal Arts and serves as the chair of the board of Directors of the Higher Education Research Institute (HEDS). She earned her PhD and MA in psychology at The Claremont Graduate University, and her BA in Psychology from Smith College.

Andrea Vasquez is a senior Engineering major receiving an emphasis in Environmental Analysis at Harvey Mudd College. Her interest in engineering design education include exploring how the design process can be adjusted to be more inclusive for a diverse set of students. In addition to her engineering research, she has been a student leader in social justice and social advocacy work. In the future, she would like to pursue work in the field of the intersection of engineering, social justice, and environmental analysis.

Joseph Sinopoli is a recent graduate of Harvey Mudd College in the field of industrial engineering. His research interests vary from achieving classroom learning outcomes via social psychology techniques, to optimization of industrial manufacturing systems, and applying operations research models to novel problems. Currently, he holds a position at Formlabs promoting the adoption of professional 3D printers in the engineering design process. Beyond this, he hopes to achieve further degrees in both engineering design and business administration.