“How?” Questions - Generalized to be open ended questions

- For questions such as “How well do the exercises cover the material?”
    - The scores/comments are varied - no inter rater reliability
- Given a question - there are examples on what a good answer should look like
    - There were a lot more lengthy comments and had a consistent score
- Whenever there is a question that is “Yes/No” in nature, the comments are limited to none
    - A solution to this is: Asking the “Yes/No” question such that it is a check box type or score type and then having a sub question that expands on their answers. This sub question needs to be only comment based without a score.
- Conclusions:
    - For questions such as “How well….”, if you’re to provide score options, remove the comment section. If comments are necessary, it is a priority to include hints on what makes a good entry.
    - Students prefer to score if the question does not explicitly ask for some sort of explanation - So the words, “Expand”, “Elucidate”, “Describe” strike a powerful chord.

“What?” Questions - Generalized to be close ended questions

- Metrics used are score and length of the answer
- When the question was positive:
    - “What are the entry’s strengths?”
        - The answers were very “flowery” and limited in length. Number of unique words was less.
        - Used words such as “enjoy”, “like” and “agree” to look for positive responses
- When the questions were negative:
    - “What are the entry’s weaknesses?”
        - The answers were very lengthy and detailed - the trend was that significantly large amounts of people want to expand on weaknesses.
        - Used words such as “not”, “disagree”, “weak” as words to look for negative responses.
        - Using negative questions led the students to believe that there *is* something wrong with the student’s paper and further beguiled them into looking deeper into the answers.
- Conclusions:
    - If a question is about how “good” something is - enforce only score. If comments are absolutely needed, have a restriction on the number of characters that can be entered.
    - If a question is about how “bad” something is - enforce compulsory comments. If scores are necessary - another constraint has to be enforced where they would allowed to score only after a certain threshold in the number of characters has been met.

        - Also, for questions with weakness related, there should be a priority on providing hints on the areas where an entry could be lacking.

“Suggest…” - Questions that ask the reviewer for feedback

- Metrics used were length and inter-rater reliability based on scores
- When the questions are too long, we got two types of responses:
    - Some students were able to use the guidance provided and gave lengthy replies, such as 208 out of 280 students gave replied greater than 300 characters.
    - Others felt lost in the lengthier questions and chose to reply in few words such as “Yes”, “No”, “adequate”.
- When the word “suggest” was explicitly mentioned:
    - The students that had felt lost earlier gave much clearer answers and were able to give strong criticism. Students were a lot more willing to expand
- Conclusion:
    - Use the keyword “suggest” in questions where instructor wants detailed feedback.
    - Have a starting question that is “Yes/No” to gauge the reaction from the student regarding the answer and then have a compulsory comment section asking for “suggestions” and then provide a scoring scale to give numerical feedback.

Mechanics - Questions dealing with grammar, structure and other standard questions

- “How correct are the mechanics?” and “Are the mechanics correct?” yielded very poor responses, where the students replied yes or no and proceeded to restate the question. For example, “Yes, the mechanics are correct.”
- For another question that had hints provided, “The write up has clean mechanics (spelling, grammar, punctuation) → The answers came pouring in. Why? *Hints.*
- Conclusion:
    - Ask the user for a score first.
    - Provide a comment box that will only switch on if the user chooses 3 or less in the scoring system. Have a word limit that is appropriate to such questions.
    - *want to conduct a study seeing if a negative spin on the question such as “What were the mechanics mistakes made” would have a different impact on students*

Ask students for their inputs on dimensions for the following questions:

- In formative assessment
    1. Check the code contributed by the author. Comment if (1) some of the functions are too long; (2) some of the code should be extracted into separate methods; (3) more comments are needed, because you have trouble following the code; (4) the code does not follow the Ruby Style Guide, or (5) you find any other code that "smells" or is not DRY.

Does the writeup explain how and why the authors did the work the way they did? If they should have used certain design principles or patterns, did they use them correctly? Comment on anything that is missing or hard to follow.

Manually test the author's work. Keep in mind of the edge cases if you see any. Do they work as intended? If you find any case for which the code does not work, please describe in the comment box in enough detail.

- In Summative Assessment:

Overall, how well was the code written. If you found problems in the code in the first round (bad names, long or complicated functions, lack of comments, bad code style, DRY problems..), have the authors improved the code accordingly?

Overall, how good is the writeup? If you found problems with the write-up (lack of explanation of the functionality, lack of explanation of how to check the work..), have the authors improved the write up accordingly?