Board 273: Exploring a Multi-dimensional Characterization of Statics Students' Questions

Kaelyn Marks, Hofstra University

Kaelyn Marks is a clinical psychology trainee within Hofstra University's Clinical Psychology Ph.D. program. She received her B.A. in Applied Psychology from University of Massachusetts Amherst, an M.S. in Psychological Science from SUNY New Paltz, and an M.A. in Clinical Psychology from Hofstra University. At Hofstra she teaches courses in psychopathology and research methods.

Dr. Saryn Goldberg, Hofstra University

Dr. Saryn R. Goldberg is an Associate Professor of Mechanical Engineering in Hofstra University's DeMatteis School of Engineering and Applied Sciences. Dr. Goldberg received her Sc.B. in Engineering with a focus on materials science from Brown University, her M.S. degree in Biomedical Engineering with a focus on biomaterials from Northwestern University, and her Ph.D. in Mechanical Engineering with a focus on biomechanics from Stanford University. At Hofstra she teaches courses in mechanical engineering and materials science. Her research in engineering education focuses on the use of student question-asking to promote metacognition.

Dr. Chris Venters, East Carolina University

Chris Venters is an Associate Professor in the Department of Engineering at East Carolina University in Greenville, North Carolina, USA. He teaches introductory courses in engineering design and mechanics and upper-level courses in fluid mechanics. He earned his Ph.D. in Engineering Education from Virginia Tech in 2014, and his research primarily focuses on conceptual understanding in engineering mechanics courses. He received his M.S. in Aerospace Engineering from Virginia Tech and his B.S. in Aerospace Engineering from North Carolina State University.

Dr. Amy M Masnick, Hofstra University

Dr. Amy Masnick is a Professor of Psychology at Hofstra University. Dr. Masnick received both her B.S. and Ph.D. in Human Development at Cornell University. At Hofstra she teaches courses in introductory psychology, research methods, cognition, and child development.

Exploring a multi-dimensional characterization of statics students' questions

Introduction

A primary goal of our DUE-funded project is to examine the quality of questions about course content asked by students enrolled in a statics course. We have developed a classroom-based intervention that provides statics students with training in the utility of question-asking and frequent opportunities to submit written questions about what they are either confused or curious about in the course. One goal of our project is to evaluate whether and how the nature and quality of student questions changes throughout the semester. To support this work, we developed a uni-axial taxonomy as a tool for characterizing students' questions [1]. This paper describes our exploration of an alternative multi-dimensional approach to characterizing student questions.

A student's ability to ask more complex questions has been identified as a way of evaluating how well a student understands a topic [2]. Therefore, the ability to ask more complex questions may be indicative of higher-level thinking about course content. Marbach-Ad and Solokove [3] used a large sample of questions generated by biology students to develop "semi-hierarchical" categories based on question sophistication [4]. Harper et al. [5] adapted this approach for classifying questions asked by physics students as part of a written reflection on their learning. Because statics is built upon physics, we used Harper et al.'s taxonomy as the basis for our own.

Previously, we shared our process for creating—and subsequently modifying—a taxonomy for use in categorizing the quality of questions students ask about statics [1]. We developed our scheme to define a higher-quality question to be one that requires or demonstrates higher-level thinking to answer – such as a question about understanding how or why something happens, or a question probing extension of knowledge to a new application – as opposed to a question that could be answered by a simple definition, or a procedural explanation of how to complete a task. Our taxonomy was approximately hierarchical, in which higher-numbered categories roughly represented metacognitively more sophisticated questions. While our modified taxonomy increased interrater reliability between faculty raters classifying student questions, a challenge remained pertaining to questions that could potentially fall into more than one category.

Consequently, in our current work, we are exploring the utility of developing a categorization system designed with the expectation that questions will fall into more than one category. This approach alleviates some challenges associated with strictly sorting questions based on the type of knowledge required to answer the question, which becomes difficult when answers require multiple or overlapping knowledge types. This approach also allows us to consider additional question features (e.g., closed- or open-ended, correct or incorrect reference to statics concepts) that may more richly evaluate question quality.

This coding approach has some similarities to the multi-dimensional taxonomy developed by Harrak et al. [6] for categorizing questions asked by first-year medicine/pharmacy students. They used a large sample of student questions to develop four independent dimensions along which they could categorize questions: question type, modality of explanation, type of explanation, and a fourth optional dimension that further defined validation/verification questions. Coding was based on identification of keywords within the questions, as one goal of their work was to enable automation of the coding process. A further goal was to see if question codes correlated with student characteristics, such as student attendance and students' grade on the final exam. They found that students with lower grades, who asked fewer questions and attended class less frequently, tended to more frequently ask "how-to" questions, as opposed to "why" questions, or questions that asked about links between concepts.

In this paper, we share our progress on developing a revised taxonomy that captures multiple dimensions of question quality and type. Specifically, we describe our process of creating the multi-dimensional taxonomy, in which some dimensions are based on our prior work on question categorization, while other dimensions attempt to capture other aspects of question structure that can affect quality based on our observation. We report preliminary findings using this new multi-dimensional taxonomy.

Method

As part of the intervention associated with this project, students are required to submit a written course-content-related question with each weekly homework assignment, resulting in a fully compliant student submitting approximately 10 questions over the course of the semester. At the same time that they submit their questions, students respond to the prompt "How comfortable do you feel with the material covered up to this point?" by selecting whether they feel "comfortable", "could use some clarification", or feel "stuck". We used data collected in Fall 2021 from one statics course with 20 students enrolled.

We first used our experience with our prior taxonomy to develop a preliminary list of question dimensions to capture the range of metacognitive complexity of student questions. Some of these dimensions were binary (e.g., open-ended or close-ended) while others required multiple levels to fully characterize (e.g., a procedural or conceptual question of low, medium or high complexity). We developed definitions for all codes and subcodes. The two statics professors who implement the intervention in their courses as part of the larger study then tried using this new scheme to code a batch of 40 student questions. The two coders then discussed any concerns or questions that came up while they were coding and refined the codes to address areas that they found problematic.

These conversations primarily resulted in adjustment to the definitions of low, medium, and high complexity subcodes within the procedural and conceptual codes. Further, we observed two categories of questions that could not be easily coded: 1) questions that contained such significant spelling or grammatical errors that we could not understand what the student was asking, and 2) questions that were readable, but for which the student had not made clear what it was that they did not understand, such that we felt the only appropriate response would be a follow-up clarification question. An example of the second category is "How does the angle affect the moment?". The coders found this category of questions difficult to code consistently, because they had to make educated guesses as to what the students were really asking about.

Code dimension	Available codes	Code definition
Related to statics content	Related, unrelated*	Question is related to course content, not course logistics
Understandable	Understandable, not understandable*	Question can be grammatically understood by the coders
Response possible	Response possible, response not possible*	Question content is clearly stated such that it could be answered, without first requiring a clarifying follow-up question
Phrased as a question	Question, not a question	Question is phrased as such
Open-ended	Open, closed	Question does not have a "yes" or "no" answer.
Procedural	Procedural (subcodes: low, med, high level), not procedural	Question is about a procedure Low: lacks statics context Med: statics context is specific to solving a particular problem or narrow circumstance High: statics context can be flexibly applied to problem-solving for a broad set of circumstances
Conceptual	Conceptual (subcodes: low, med, high level), not conceptual	Question is about a concept Low: seeks definition of a term Med: seeks an explanation involving statics, that is not a single definition and is not indicative of high-level^ thinking High: seeks an explanation that extends beyond statics or is indicative of high-level^ thinking
Includes only correct statements related to statics content	Correct, not incorrect	Question does not include a false or incorrect statement about statics content
Strategic question	Strategic, not strategic	Question is about how to plan or optimize approach to problem-solving
Wonderment question	Wonderment, not wonderment	Question explores boundaries of application of statics content, such as "What would happen if…"

Table 1. Current coding scheme. Codes marked with a * indicates that this code would result in the end of the coding process, e.g. a question that was unrelated to statics content would not be coded in any subsequent categories. ^We recognize that the definition of high-level needs to be more explicitly defined. A next step in the code-refining process is to look at the questions that are coded consistently as high-level conceptual and clarify what characteristics they share.

Agreeing that questions that fit into either of these two problematic categories would not be coded further, the coders used the refined code set on another group of 40 student questions, and then met to discuss concerns. This conversation resulted in minimal further refinements. The coders used the resulting scheme (Table 1) to code the entire batch of 160 questions. The final coding scheme included 10 dimensions, each of which were binary except for procedural and conceptual, which receive subcodes of low, med, or high.

Results

A total of 188 student responses were coded. Of these, 30 (16%) were determined to be unrelated to statics content. In most cases, these represented cases where students asked about course logistics related to grading or exam coverage. For example, "What are going to be the main focuses on the next exam, like material-wise?"

Table 2 provides a summary of the number of questions coded in each category and the agreement between coders. Of the 158 questions that were related to course content, 145 (92%) were coded "understandable" by both coders. Further, both coders felt that they would be able to provide responses to 107 (74%) of the 145 understandable questions. This became the pool of questions that were fully coded. There were 12 questions that both coders could understand but neither coder could provide a response to without more information from the student. Cases in which the student didn't phrase their response as a question were still coded if it was still clear what the question form would be. For example, "I don't understand what the word moment means."

Code dimension	**Codes agreed upon / questions coded (% match)**
Related to statics content	158 / 158 (100%)
Understandable	145 / 158 (92%)
Response possible	119 / 145 (82%)
Phrased as a question	107 / 107 (100%)
Open-ended	103 / 107 (96%)
Procedural	91 / 107 (85%)
Conceptual	95 / 107 (89%)
Includes only correct statements related to statics content	103 / 107 (96%)
Strategic question	96 / 107 (90%)
Wonderment question	97 / 107 (91%)

Table 2. Number of questions coded in each category and the agreement between coders.

We note that of the 13 responses that were deemed "not understandable" by a single coder, 11 were deemed "not answerable" by the same coder, meaning that they would have not been coded further anyway. There were 26 disagreements between coders when determining if a response could be provided to an understandable question. This comprises 18% of the 145 understandable questions; however, in all but one case, the coder who felt that a response could not be provided was not the instructor of the course. This is likely the result of not being privy to certain contextual features of the classroom setting.

For the 107 questions that were fully coded, the agreement between coders was generally good. Dimensions that featured binary choices all had agreement greater than 90%. Procedural and conceptual dimensions, which each had one primary level and three subcode levels, saw agreement at 85% and 89%, respectively, at the level of the primary code.

Additional analyses were conducted to view the frequency of question types among students of varying academic achievement levels. Students were put into categories based on the mean final exam score. The final exam was used as a proxy for a student's academic achievement level because it was a cumulative exam, so the grade on the exam would reflect a student's overall level of understanding of the statics material covered in the course. The mean exam score in the course was 62.46% with a SD of 21.58. Students who fell within 1 SD of the mean were considered to display an average level of academic achievement. Students above and below 1 SD of the mean score were considered to display a level superior to or below the average achievement level. Thus, scores below 40.88% and above 84.04% were within these categories. Five students were within the below achievement category, resulting in a total of 27 questions, and an average of 5.4 questions each. Eleven students were within the average range, resulting in a total of 71 questions, and an average of 6.5 questions each. Three students performed in the superior range, for a total of 25 questions, and therefore 8.3 questions each. The variation in average questions asked per student is in part due to the higher frequency of questions which were not answerable or relevant among students in the average and below average categories.

Table 3 provides details as to how question type varied with student achievement level. Notably, within the procedural category, students in the low and average achievement levels primarily asked low-level procedural questions. In contrast, students in the superior range predominantly asked high-level procedural questions.

Students also indicated their current level of comfort with the material in the course for each question submitted throughout the course. As seen in Table 4, when categorizing the students' status of comfort in the course by achievement level, the primary status of students in the below and average levels of achievement was that they expressed needing clarification. The primary status of students in the superior level of achievement was that they felt comfortable.

Finally, we were interested in what types of questions students asked when they indicated that they were stuck. The status of being stuck accompanied 18 questions in the sample. Results are summarized in Table 5. Notably, low-level procedural questions were the most commonly asked questions for these students. This led us to ask what qualities are shared with high-level procedural questions. Twenty-five questions were within the high-level procedural category. Analysis of concurrent categories with the high-level procedural questions showed that none of the questions in this category shared the features of a conceptual question. The primary status of students who asked these questions was that they felt they needed clarification (Table 6).

Achievement level	Open	Correct	Strategic	Wonderment	Complexity level	Proced.	Concep.
Below	52%	0%	15%	22%			
					Low	41%	4%
					Med	11%	22%
					High	15%	4%
Average	40%	8%	6%	6%			
					Low	32%	56%
					Med	10%	30%
					High	15%	8%
Superior	60%	4%	20%	3%			
					Low	20%	8%
					Med	0%	32%
					High	40%	4%

Table 3. Question type by student level of achievement.

Achievement level	Status	
Below	*Comfortable*	30%
	Need clarification	59%
	Stuck	11%
Average	*Comfortable*	24%
	Need clarification	58%
	Stuck	18%
Superior	*Comfortable*	52%
	Need clarification	40%
	Stuck	8%

Table 4. Comfort status among students across achievement levels.

Open	Correct	Strategic	Wonderment	Level	Proced.	Concept.
72%	1%	6%	12%			
				Low	44%	6%
				Med	6%	22%
				High	17%	6%

Table 5. Types of questions students asked when stuck.

Open	Strategic	Wonderment	Level	Concept.	Status	
68%	24%	4%				
			Low	0%	*Comfort*	32%
			Med	0%	*Need clarification*	56%
			High	0%	*Stuck*	12%

Table 6. Concurrent categories with high-level procedural questions.

Discussion

While our results at this point are preliminary due to the relatively small set of questions assessed and further work needed to refine definitions for the procedural and conceptual subcodes, we can make some initial observations, both about the ease of use of the multi-dimensional taxonomy and what it may reveal about student questions.

Both faculty raters found this taxonomy easier to use than our original, semi-hierarchical taxonomy. The coding process was faster and felt more natural. To use the previous taxonomy, we found that it was necessary to include a list of how we had agreed to code a range of different question cases. This was necessary, in part, because the five available levels did not adequately capture the multi-faceted nature of many questions that students asked. The current multi-dimensional approach is better able to capture this complexity, and as a result, we have confidence that the coding definitions alone will be enough to enable consistent coding, as seen in the high levels of agreement between coders. This scheme included codes for questions that could not be understood or answered, eliminating them from further coding, which also contributed to it being easier to use.

In our many semesters of working with student questions, we have observed that higher-performing students ask questions that support their being able to apply the skills they are learning to broadly solve statics problems, while lower-level students are more focused on getting the correct answer in a specific problem or circumstance. The current taxonomy's characterization of procedural and conceptual questions as being low-, medium- and high-level questions appears to support this observation. The coding scheme defines low-level procedural questions as asking about procedures without a connection to a statics context (i.e. How do I take the cross product?), medium-level procedural questions as providing a statics context that is specific to a particular problem or situation (i.e. For question 3 on the HW, should I include the force from the wall?), and high-level procedural questions as asking about how to broadly solve classes of statics problems (i.e. When solving for moments in a non-planar system do we always start with Mnet = 0?).

We found that high-performing students asked high-level procedural questions more often than low- or medium-performing students. Further, high-level procedural questions were infrequently asked by students who described their current status as "stuck". Instead, stuck students were more commonly asking low-level procedural questions. This is consistent with the idea that high-performing students are integrating their understanding into skills that will serve their problem-solving across contexts. While conversely, low-performing students are less aware of what they need to know to become more proficient problem solvers and are not as readily making connections between the specific problem they are solving and statics concepts more broadly.

Our work with this new characterization approach is still preliminary. To further validate the approach, we need to apply our new coding scheme to a larger data set while looking at other question types that challenge inter-rater reliability. We may find new question categorizations which show poor reliability, which could lead to further refinement of our coding definitions. We plan to explore adding additional dimensions via employing an inductive coding approach to identify new themes within student questions. We need to repeat the analyses reported here with questions across multiple semesters and institutions to confirm their robustness. Future work will also include using the taxonomy to look at how student questions change over the course of the semester and identifying whether specific cohorts of students (i.e. high-performing students or students who show improvement in performance over the semester) ask questions with identifiably different characteristics.

References

[1] S. Goldberg, C. Venters, and A. Masnick, “Refining a taxonomy for categorizing the quality of engineering student questions,” in *2021 ASEE Annual Conference and Exposition Proceedings,* virtual, July 2021.

[2] Y. J. Dori and O. Herscovitz, “Question-posing capability as an alternative evaluation method: Analysis of an environmental case study,” *J. Res. Sci. Teach.*, vol. 36, no. 4, pp. 411–430, Apr. 1999, doi: 10.1002/(SICI)1098-2736(199904)36:4<411::AID-TEA2>3.0.CO;2-E

[3] G. Marbach-Ad and P. G. Sokolove, “Can undergraduate biology students learn to ask higher level questions?” *J. Res. Sci. Teach.*, vol. 37, no. 8, pp. 854–870, Oct. 2000, doi: 10.1002/1098-2736(200010)37:8<854::AID-TEA6>3.0.CO;2-5.

[4] G. Marbach-Ad and P. G. Sokolove, “Good science begins with good questions,” *Journal of College Science Teaching*, vol. 30, no. 3, pp. 192–195, 2000.

[5] K.A. Harper, E. Etkina, and Y. Lin, “Encouraging and analyzing student questions in a large physics course: Meaningful patterns for instructors - Harper - 2003 - Journal of Research in Science Teaching - Wiley Online Library,” *Journal of Research in Science Teaching*, vol. 40, pp. 776–791, 2003.

[6] F. Harrak, F. Bouchet, and V. Luengo, “Identifying relationships between students’ questions type and their behavior,” in *10th International Conference on Educational Data Mining*, Wuhan, China: IEDMS, Jun. 2017. Accessed: Feb. 05, 2024. [Online]. Available: https://hal.science/hal-01525285