2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #34062

Refining a Taxonomy for Categorizing the Quality of Engineering Student Questions

Dr. Saryn R. 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 Assistant 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 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, cognitive psychology, and child development. Dr. Masnick is interested in conceptual development, reasoning about science and number in children and adults, and in science and engineering education.

Collaborative Research: Refining a Taxonomy for Categorizing the Quality of Engineering Student Questions

The ability to identify one's own confusion and to ask a question that resolves it is an essential metacognitive skill that supports self-regulation [1]. Yet, while students receive substantial training in how to *answer* questions, little classroom time is spent training students how to *ask* good questions. The study presented here is from a pilot conducted in preparation for a larger study funded through NSF-DUE that investigates if training students to ask better questions, and giving them practice and feedback on writing those questions, affects other important STEM learning outcomes.

One challenge in research around question-asking is defining what constitutes a good question, as there are many ways in which a question may be characterized. Researchers have used a range of approaches to categorize students' questions, varying in complexity depending on the context in which student questions were being solicited (e.g., [2], [3]). Marbach-Ad and Solokove [4] used a large sample of questions generated by biology students to develop a six-level, "semi-hierarchical" taxonomy based on question sophistication. Encouragingly, their work also shows that students are able to pose more high-quality questions after being instructed in the taxonomy for classifying the quality of their questions [5]. This approach has also been adapted for classifying questions asked by physics students as part of a written reflection on their learning [6].

Along with explanatory question taxonomies, question-asking can be additionally scaffolded by providing students with question stems [7]. Specifically, King [8] investigated the use of question stems as a guide for college students to pose higher-level questions. Importantly, this was an example of "informed" strategy training [9]: the participants were provided with the reason for using the strategy and how it could potentially improve learning. This approach is consistent with the more recent finding that perceptions of task value influence the learning process, including in the specific context of question generation [10].

To pilot the intervention to be included in our larger study, at the end of each class period during a semester-long engineering statics course, students were prompted to write and categorize a question that they believed would help them clarify their current point of greatest confusion. A sample of these questions were selected to be categorized, answered, and made available to the class online as a learning resource. Through regular practice writing and categorizing such questions, we aimed to improve students' abilities to ask questions that require higher-level thinking. For this pilot study, we investigated the utility of the question taxonomy in rating actual student questions.

Informed by prior literature, we created and deployed training materials prior to the intervention to raise student awareness of the utility of asking high-quality questions. We also taught them how to use a question classification taxonomy that includes built-in scaffolding in the form of example questions and question stems. Because statics is built upon physics, the classification scheme presented by Harper et al. [6] served as the basis for the taxonomy we created for our work. Our taxonomy specified five question types, as seen in Table 1. We developed this scheme to define a better or higher-quality question to be one that requires or demonstrates higher-level

thinking. For example, a high-level question would be 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 perform a specific task. In the case of a classroom intervention like the one we discuss here, in which students pose questions in response to a given prompt, the questioning context is well-defined, and therefore we chose a simple uniaxial scale for assessing questions, with the belief that it could capture the range of expected responses.

Table 1. The original question classification taxonomy used in pilot data collection in Fall 2019.

Question Type	Question Description	Example Questions and Question Stems
1	Question is unspecific.	What is going on? How do I solve the problem?
2	Question is about a definition.	What is…? What is the difference between…and...?
3	Question is about how to do something.	How do I…?
4	Question is about understanding how or why something happens.	Why…? How…? How does...affect…? How is...related to…?
5	Question extends knowledge to a different circumstance.	Could...? What would happen if…?

In Fall 2019, we collected data from 35 students in courses at two institutions. Over the course of the semester, students had the opportunity to write and categorize twenty of their own questions. After the semester, faculty from each institution categorized student questions using the taxonomy to assess the appropriateness of the taxonomy, whether students used it accurately, and the consistency between faculty raters.

Analysis of the pilot data indicates three issues to be addressed:

1. Student compliance in writing and categorizing their questions varied. As originally designed, our intervention prompted students to write a question at the end of every class period. This would result in a fully compliant student writing 20 questions in about 10 weeks (as the questioning protocol was not started until the 3rd week of the semester and was not used in the final week of the semester). We received some feedback from students that this frequency felt burdensome, and that this diminished their motivation to engage with the intervention. At one institution student compliance averaged 72%, and at the second institution it averaged 64%.
2. Some students had difficulty correctly coding their questions using the taxonomy. When student question categorization was compared to faculty categorization of the same questions, students were only found to be able to code their own questions correctly 66% of the time. This discrepancy suggests that students either didn't understand the

categories, or that the categories did not accurately span the range of questions that students ask.

3. Some student questions could not be clearly characterized using the taxonomy, even for faculty raters. As part of the analysis of our initial pilot data, we noted that we found many questions difficult to categorize using our existing taxonomy, which contributed to inaccurate student categorizations and poor consistency between faculty raters. This made it clear that the taxonomy needed to be adjusted so that it more effectively and clearly spanned the questions asked by students.

The ways we will address each of these issues are elaborated on below as we describe the changes we made to the prompt, to student training and to the taxonomy itself.

Modifications to the prompt

To address low student compliance with the intervention we decided to prompt students to write questions less frequently. In Spring 2020 - the semester that was interrupted by the global pandemic - one faculty member implemented the intervention and modified it to require students to write just one question per week, with the option to write more if they chose. While the intervention was only undertaken for 4 weeks, compliance increased to 90%, with 18 out of 29 students choosing to write more than one question per week. We are therefore hopeful that this modification will result in more consistent student participation.

In designing the question prompt, we recognized that some students may find that they are not confused about anything and therefore may claim that "they don't have a question." To give these students an opportunity to still engage with the prompt, we asked them to write a Type 5 question that extended their understanding to new contexts. As a result, we think we had more students writing questions of this type that were not very well-conceived, likely because doing so was a way of avoiding grappling with the material at hand. These questions were often difficult for faculty to categorize accurately, as they were often ill-posed. To avoid this going forward, we have modified the prompt to ask students who are not confused to write a question that they think a student who has a good understanding of the material should be able to answer - and then answer it. We are hoping that this added requirement will incentivize more students to think about what genuine questions they might have and reduce the number of frivolous questions.

Modifications to student training

Reducing the amount of time in class that students spend writing questions also supports our being able to increase the amount of class time engaged in training students in the use of and utility of the question categorization taxonomy. This increase in training time was motivated by our students' poor ability to use the taxonomy to categorize their questions correctly. We also made changes to the training to help students better understand how to use the taxonomy.

In Fall 2019, training consisted of students completing a short reading that exposed them to the ideas of procedural and conceptual learning, emphasizing the importance of both. These ideas were then discussed in class in relation to the importance of asking questions to support conceptual understanding. The taxonomy was introduced in class and students were given some questions to practice categorizing. For the next two class meetings, students were walked

through the process of writing a question at the end of class. Students were invited to share their questions out loud and correct categorization was discussed.

In light of students' low accuracy in using the taxonomy, we redesigned the training to offer students more frequent opportunities to practice using the taxonomy and get feedback on its use. We are piloting these changes during the current, Spring 2021 semester. The training has been expanded to consist of four approximately 10-minute in-class sessions spread out over two weeks, plus additional homework on the topic assigned outside of class, which is described in more detail in the paragraph below. Students are also given feedback on ratings of their own questions throughout the semester as faculty select questions to answer and post for the class.

In the first session of the revised training, the faculty member leads a discussion asking students to think about how they recognize when they are confused about something and what their options are to help them to resolve confusions that they have, highlighting that most approaches to resolving a confusion require asking a question. For homework, students are asked to complete the previously described short reading about procedural and conceptual learning, which is then discussed briefly in the subsequent class meeting (second training session). The third in-class session begins with a discussion of what defines a "good" question and introduces the question taxonomy. As part of this discussion, the faculty member presents two sets of questions about topics that the students already understand: trigonometry and vectors. Each question set contains one question from each type in the taxonomy. The faculty member also leads a discussion of how to categorize these questions, which then become example questions that are a part of the taxonomy table presented to students as part of the question prompt. As homework, students are asked to write five questions - one of each type - about any topic of their choice. The faculty member provides feedback to each student on how well these questions are categorized. Students are assigned a section of the textbook to read about the topic of moments after being covered in class. They are asked to write and categorize two questions that they have about moments. In the fourth and final in-class session, the faculty member shares (anonymously) a selection of these questions and discusses their correct categorization. She then introduces the prompt and the logistics of how it will be used as part of the course.

While data collection following this new training approach is still underway, the students seemed to do well with the initial question categorization tasks, giving us some confidence that the changes we have made will improve students' ability to accurately use the taxonomy to correctly categorize their questions.

Modifications to the taxonomy

We began our process for modifying the taxonomy to more accurately span the questions asked by students by discussing the questions that the two faculty members categorized differently. These conversations led to the following realizations and changes and resulted in the modified taxonomy shown in Table 2:

1. While question stems are an important and appropriate scaffolding tool to support students in writing their own questions, aligning stems with specific question types led some students to mis-categorize questions simply because they started with a particular

word. To address this, we have separated the question stems as a separate resource shared with students and replaced them within the taxonomy chart with example questions from topics that students are familiar with (and which were discussed in class as part of the training).

2. While students sometimes do write unspecific questions, having these listed as a type within the question taxonomy may encourage this type of question-writing, which we don't see as very productive for students. Faculty will retain this question type (Type 0 in Table 2) when categorizing student questions, but it will not be shared with students as part of the taxonomy.
3. Looking at the questions students wrote, we saw that there were two classes of "how to do something" questions: one type was a very basic procedural question and another type that was more strategic, showing some student understanding of what to do and aimed at understanding nuance. To capture this distinction, we introduced a new question type (Type 3 in the modified taxonomy in Table 2) between the "procedural" and "conceptual" types of questions, which we are calling a "strategic" question.

The faculty raters' difficulty in categorizing student questions using the original taxonomy (Table 1) led to low inter-rater consistency. This motivated our documenting a shared understanding for how we would categorize commonly asked sub-categories of questions.

We noted that the questions that we most commonly rated differently were questions on the boundary between categories (i.e., one faculty member would rate it as a 3 and the other would rate it as a 4) and category 5 questions. Discussing details about our categorization choices for questions that we rated differently contributed to some of the changes made to the taxonomy, as discussed in the previous section. It also motivated our change to the prompt to no longer ask students with no confusion to write a Type-5 question. These conversations also enabled us to generate internal guidelines to use when categorizing questions, such that we had an agreed upon way of categorizing question types that often led to mismatch. For example, we decided that questions asking "What is the difference between X and Y?" would be categorized as Type 4 questions, as opposed to Type 1 questions, because what they are asking about goes beyond understanding simple definitions and is instead making connections between concepts.

We went through multiple rounds of making modifications to the taxonomy, which consisted of both faculty members rating another batch of questions with the most recent version of the taxonomy and discussing any questions with mismatched ratings. Initially the faculty agreement on ratings was approximately 65%. With each iteration of modifications, the faculty members felt more confident in their categorization choices, but the overall agreement rate did not significantly improve. After three rounds of this, using different batches of student questions from Fall 2019, the largest category of inconsistent question ratings occurs when one faculty member rated the questions of the other faculty member's student a Type 0 due to lack of context, while the faculty member whose student wrote the question was able to categorize it. These conflicts were easily resolved with a conversation between faculty raters, but we also hope they will be minimized by giving students the explicit instruction to write questions that stand on their own. Specifically, we will now ask students to word their questions in a way that they could

Table 2. The modified question classification taxonomy, including example questions and separate accompanying list of question stems, used in Spring 2021.

Question Type	Question Description	Example Questions
0*	**Question is unspecific.**	** This category is only known to faculty raters. It provides a consistent categorization for questions that do not fit other categories due to lack of specificity, i.e. "How do I get started?"*
1	**Question is about a definition** - a "foundational" question	*How do you define a "right triangle"?* *What do you mean by vector "magnitude"?*
2	**Question is about how to do something** – a "procedural" question	*How do I determine the third angle in a triangle if given values for two angles?* *How do I make a unit vector that points from a given point A to a given point B?*
3	**Question is about how to do something if conditions of the problem changed** – a "strategic" question	*Should I always use cosine to solve for the x-component of a vector?* *Can I use the same unit vector to create a position vector and a force vector?*
4	**Question is about understanding how or why something happens** – a "conceptual" question	*Why do two different angle values correspond to the same value of sin(theta)?* *Why doesn't a unit vector have any units?*
5	**Question extends knowledge to a new circumstance beyond that of the problems solved in class** – a "wonderment" question	*Do equations like the Law of Sines and Law of Cosines exist that allow for you to solve for side lengths or angles of quadrilaterals?* *Are there such things as 4-dimensional vectors?*

Question Stems
What is…? Why does…? How does…? What is the meaning of…? What is the difference between…and…? How does...affect…? How is...related to…? Could...? What would happen if…?

still understand them if looking back at them a month later. If we exclude this category of mismatch between faculty raters, since it arises primarily due to the fact that the two faculty raters teach separate courses, faculty agreement on ratings improved to approximately 80%.

Next steps

While the current semester, Spring 2021, is impacted by the global pandemic, with most teaching happening online, we are still using the opportunity to continue to pilot changes to our intervention in one section of statics at one institution. Students are prompted to write and categorize questions through an online Learning Management System (LMS) whenever they submit a homework assignment, approximately once per week. The faculty member is using the LMS to provide feedback on student categorization and answer student questions.

As discussed above, we are piloting the new training methodology and refined taxonomy during the current semester and will be able to report updated data at the ASEE conference. While teaching methods are still impacted by the pandemic, we still see this as a valuable opportunity to learn what works and what doesn't before (we hope) a return to in-class teaching in the fall, when we will be able to start collecting the full complement of data associated with our broader study activities.

The current online teaching format means that all lectures are recorded, which will enable the full team to review the training methodologies together to make refinements for the fall. Further, all student questions and categorizations are collected online, which greatly streamlines data capture.

We plan to apply the same analyses described here to the new set of student questions collected during the Spring 2021 semester, such that we can make final adjustments to the taxonomy before using it as part of our full intervention in Fall 2021.

References

[1] P. H. Winne, "A perspective on state-of-the-art research on self-regulated learning," *Instr. Sci.*, vol. 33, no. 5, pp. 559–565, Nov. 2005, doi: 10.1007/s11251-005-1280-9.

[2] A. C. Graesser and N. K. Person, "Question asking during tutoring," *Am. Educ. Res. J.*, vol. 31, no. 1, pp. 104–137, 1994, doi: 10.2307/1163269.

[3] H. Pedrosa De Jesus, J. J. C. Teixeira-Dias, and M. Watts, "Questions of chemistry," *Int. J. Sci. Educ.*, vol. 25, no. 8, pp. 1015–1034, Aug. 2003, doi: 10.1080/09500690305022.

[4] G. Marbach-Ad and P. G. Sokolove, "Good science begins with good questions: answering the need for high-level questions in science," *J. Coll. Sci. Teach.*, vol. 30, no. 3, pp. 192–195, 2000.

[5] 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, 2000, doi: 10.1002/1098-2736(200010)37:8<854::AID-TEA6>3.0.CO;2-5.

[6] K. A. Harper, E. Etkina, and Y. Lin, "Encouraging and analyzing student questions in a large physics course: Meaningful patterns for instructors," *J. Res. Sci. Teach.*, vol. 40, no. 8, pp. 776–791, 2003, doi: 10.1002/tea.10111.

[7] F.-Y. Yu, H.-C. Tsai, and H.-L. Wu, “Effects of online procedural scaffolds and the timing of scaffolding provision on elementary Taiwanese students’ question-generation in a science class,” *Australas. J. Educ. Technol.*, vol. 29, no. 3, pp. 416–433, Jun. 2013, doi: 10.14742/ajet.197.
[8] A. King, “Effects of self-questioning training on college students’ comprehension of lectures,” *Contemp. Educ. Psychol.*, vol. 14, no. 4, pp. 366–381, 1989, doi: 10.1016/0361-476X(89)90022-2.
[9] A. L. Brown and A. S. Palincsar, “Inducing Strategic Learning from Texts by Means of Informed, Self-Control Training. Technical Report No. 262,” Illinois Univ., Urbana. Center for the Study of Reading & Bolt, Beranek and Newman, Inc., Cambridge, MA, 1982.
[10] F.-Y. Yu and C.-P. Wu, “Student question-generation: The learning processes involved and their relationships with students’ perceived value,” *J. Res. Educ. Sci.*, vol. 57, no. 4, pp. 135–162, 2012.