Work in Progress: Context Matters: A Comparative Study of Results of Common Concept Questions in Statics at Several Diverse Institutions

Abstract. Four institutions collaborated to deploy a set of four common concept questions from the Concept Warehouse. In responding to the concept questions, student not only directly answered the question but were also asked to provide written explanations of their reasoning, ratings of their confidence with their answer, and ratings of their perceptions on the question effectiveness to help them learn. The research questions addressed by this preliminary study were: *How do faculty at different institutions employ concept-based instruction?* and *How do students in different contexts respond to concept questions?* The study identified three modes of deployment, two modes used concept questions before or during introduction of a new topic and one mode used them sometime after concept introduction. In all cases, the grading basis was "low stakes". Concerning response to concept questions, preliminary results show that students identifying as female express lower confidence in their answers compared with students identifying as male whether or not they outperformed male student. This finding was consistent across all institutional settings. Students also had difficulty providing essential and correct explanations for phenomena, even when they provide the correct response, regardless of institutional setting or gender identity. Finally, there also appears to be no correlation between performance, confidence, and question effectiveness.

1. Introduction

The Concept Warehouse ("CW") is an interactive repository for deployment of several thousand concept questions called "ConcepTests" that range over several topics in engineering, including approximately 700 in engineering mechanics. Concept questions are qualitative in nature and are designed to elicit patterns of thought that complement those required for procedural questions. Concept questions may be used in a variety of modalities, online or offline, in-class or out of class, and with response time allocated to be "immediate" (say 2-5 minutes during class) or "protracted" (say several hours or days as a preparatory exercise or homework).

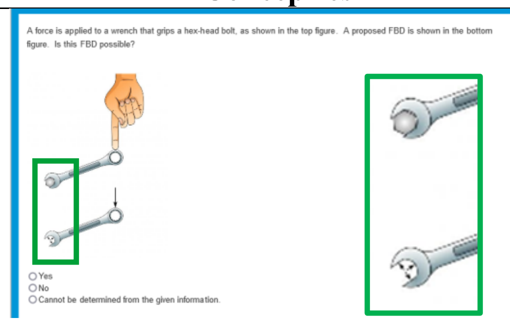
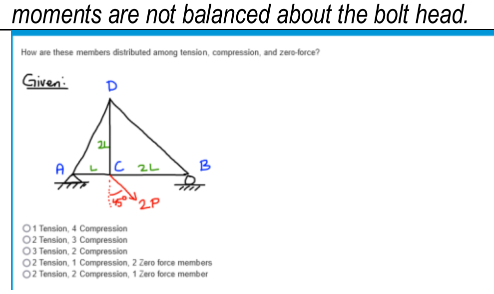
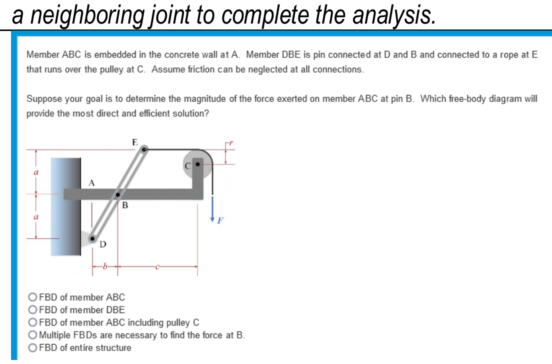
Four Statics questions from the Concept Warehouse were selected to be used by a group of collaborating faculty at four institutions during Fall 2021 (and continuing into Spring 2022). These questions were selected based on their range over core topics (rigid body equilibrium, trusses, frames, and friction) and also due to their established usage patterns. The participating faculty assigned these questions to their students at the point and in the modality that they deemed to be appropriate.

This study comes from a community of practice with the overarching goal of improving student success in engineering mechanics courses. The guiding questions behind this particular study are *How do faculty at different institutions employ concept-based instruction?* and *How do students in different contexts respond to concept questions?* The data presented here is preliminary and is being reviewed to determine emerging patterns. Moreover, this work is part of a broader project with other researchers and colleagues who are examining contextual factors to understand both faculty and student experiences and attitudes toward concept-based instruction using the CW. A survey is currently being deployed to students to gain further data of this type, but the results are not available for this study.

2. Description of Common Questions Study

The four participating institutions in this study include the University of Puerto Rico, Mayagüez, a public, bilingual, Hispanic-serving institution; Whatcom Community College, a small community college in the state of Washington; Elizabethtown College, a small private college in Pennsylvania; and Oklahoma State University, a large public research university.

Table 1. The four common questions used in this study.

ID	Topic and Text	ConcepTest
4550	**Rigid Body Equilibrium** "A force is applied to a wrench that grips a hex-head bolt, as shown in the top figure. A proposed FBD is shown in the bottom figure. Is this FBD possible?" • Yes • No • Cannot be determined from the given information	 *The detail of the key region is provided for clarity, but does not appear in the actual question. The intention is for students to focus on the FBD and realize that moments are not balanced about the bolt head.*
4606/ 4756	**Trusses** "How are these members distributed among tension, compression, and zero-force?" • 1 Tension, 4 Compression • 2 Tension, 3 Compression • 3 Tension, 2 Compression • 2 Tension, 1 Compression, 2 Zero-force members • 2 Tension, 2 Compression, 1 Zero-force member	 *The intention is for students to draw mental or actual FBDs of various joint, make qualitative determinations about the modality of the member, and then continue to a neighboring joint to complete the analysis.*
5134	**Frames and Machines** "Member ABC is embedded in the concrete wall at A. Member DBE is pin connected at D and B is connected to a rope at E that runs over the pulley at C. Assume that friction can be neglected at all connections. Suppose your goal is to determine the magnitude of the force exerted on member ABC at pin B. Which free-body diagram will provide the most direct and efficient solution?" • FBD of member ABC • FBD of member DBE • FBD of member ABC including pulley C • Multiple FBDs are necessary … • FBD of entire structure	 *The intention is for students to draw mental or actual FBDs of various members and determine which one provides a solvable set of equations that includes the pin force at B.*

4497	**Friction (Particle Equilibrium)** "You are holding a box of books with flat hands. If you press harder, what happens to the friction force applied by your hands onto the sides of the box?" • It increases • It remains the same • It decreases • Not enough information to determine	 *The intention is for students confront a simple situation in which the common law "$F = \mu N$" does not apply, and to realize the importance of drawing a simple FBD and applying equilibrium.*

The four common questions are summarized in Table 1 by providing a screenshot of the problem as it would appear to a student in the Concept Warehouse. These problems can be accessed by requesting an instructor account in the CW and then searching for the ID number provided (https://newjimi.cce.oregonstate.edu/concept_warehouse/). As discussed in [1], each question has a set of intended goals, and these are summarized in the table. In general, these goals are commensurate with the nature of concept questions in that qualitative reasoning is prioritized above procedural calculations.

One primary objective of concept-based teaching and learning is not only to engage students with the question itself, but also to engage students in explaining their reasoning and thinking about how sure they are about their answers [2]. Therefore, as illustrated in Figure 1, the CW has an option to allow students to explain their reasoning, and to express their level of confidence in their answer. For the purposes of data tabulation, the confidence responses are recorded as integers from 1 to 5, ranging from "substantially unsure" = 1 to "substantially confident" = 5.

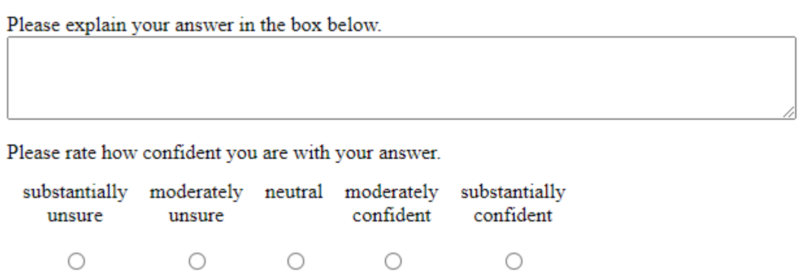


Figure 1. Screenshot of the section where students provide explanations and confidence ratings corresponding to their responses.

Before proceeding, it is important to note that some concept questions in the CW have been intentionally designed to be ill-posed, such as, for example, to have more than one "defensible solution". Epistemologically, the purpose of this is to counter-balance the dominant notion that "all engineering problems have a single correct answer", and thus have a "meta"-goal to help students renormalize their conceptions of what it means to solve an engineering problem, even when it appears to be "straight forward". However, even when consciously desired, developing and deploying such problems proves to be difficult, for reasons ranging from requiring more effort to create and score problems, to having time for discussion in institutional systems with rigid curricula that never accounted for this kind of interactive exploration in the first place. Therefore, to date, the vast majority of questions in the CW are "well-posed".

For this reason, and also for simplicity of data analysis, the questions in this study are well posed. That is, each of these questions has a single "correct" answer and student reasoning can be nominally judged to be "correct" or "incorrect". However, the authors acknowledge that within "incorrect" reasons lie a host of nuanced ideas that students are attempting to express, many of which have some degree of "correctness" that is useful for class discussion. For this reason, in a classroom situation, the authors usually avoid declaring an idea "correct" or "incorrect", but rather frame responses as "useful", "helpful", or "reasonable", so as to encourage student participation and to help make connections with their initial reasoning to skilled mechanical reasoning.

Recently, additional meta-questions were added to the CW to further understand student impressions on "question effectiveness", as measured by the degrees to which the question was understood ("UQ") and made them "think deeply" ("TD"). This portion of the platform is illustrated in Figure 2. For the purposes of data tabulation, the confidence responses are recorded as integers from 1 to 5, ranging from "strongly disagree" = 1 to "strongly agree" = 5.

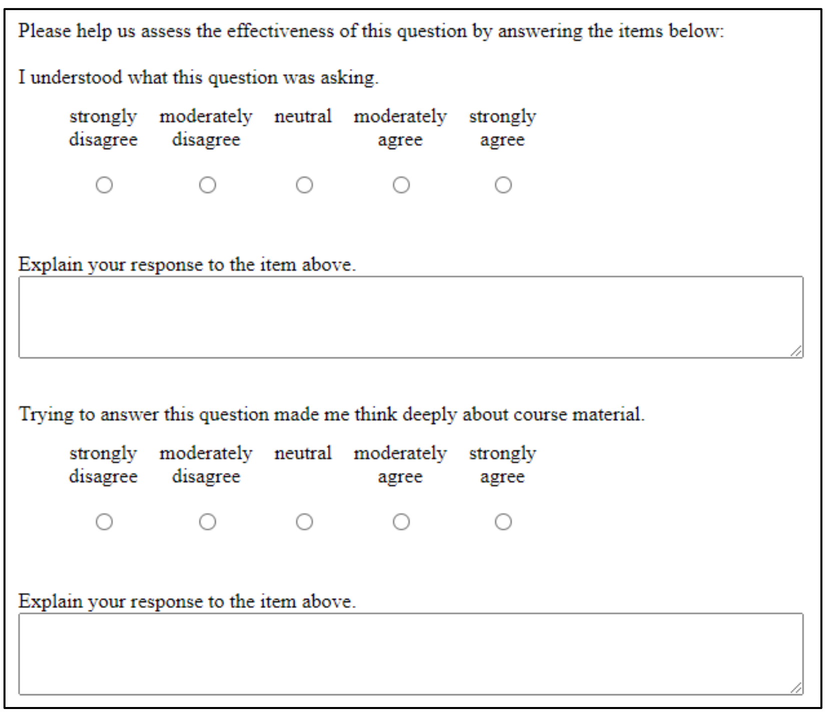


Figure 2. Screenshot of the section where students respond to question effectiveness.

Another feature of concept-based teaching and learning is that is can be employed flexibly according to the instructors' and students' dispositions and needs. Table 2 provides details of the institutional profiles and the modality of how the questions were deployed. The diversity of modality is noteworthy, and highlights the usefulness of the CW as a tool that can be employed flexibly. Yet, what is common is that all instructors in this study are using the CW for "low stakes" assignments or activities.

Table 2. Modality of Deployment by Institution

Institution	Modality of Deployment
University of Puerto Rico, Mayagüez (public, bilingual, Hispanic-serving research university)	ConcepTests were assigned as required homework, and were expected to be completed as part of the initial learning of the topic. The grade was based on completeness, not on correctness. Some discussion occurred in class before all students had completed the assignment, meaning that some students' scores reflect their diligence in class, rather than their direct approach to the question.
Whatcom Community College (small community college)	ConcepTests deployed primarily in two modalities throughout the course. CTs 4550, 4756, and 4497 were three of dozens of CTs used regularly during peer instruction in class. CT 5134 was one of a series assigned for weekly homework in a pre/post implementation at the start and end of each week of material. For those assigned as homework, students earned full credit for selecting an answer and entering an explanation regardless of accuracy.
Elizabethtown College (small private college)	The four ConceptTests were administered during regular class time, in a single session. Questions 4550, 4756, and 5134 were presented after the corresponding content was covered. However, question 4497 was presented right before introducing the corresponding learning module. Students were given a limited time to individually complete the questions in the classroom. Then the instructor led a discussion on the questions and answers.
Oklahoma State University (large public research university)	ConceptTests were assigned as homework, with a small amount of bonus points being awarded for completeness.

3. Results

Data are collected in the CW platform and then can be downloaded on demand for analysis. The data collected here corresponds to student responses at the four institutions during Fall 2021. Due to the 'well posed' nature of the questions, the data includes a designation of the correctness of each response, from which the "percent correct" is easily determined. To assess reasoning, each instructor manually reviewed the written explanations of the responding students and assigned a "1" or "0" to determine if the reasoning is "correct" or "incorrect", and therefore not attempting to provide a more nuanced scoring to allow for the variety of "helpful" or "useful" responses that might be embedded (see discussion in Section 2). From this, the instructors further assigned a score of "C+C" (2 pts) if both the response and reasoning were "correct", "MM" (1 pt.) if the reasoning is mismatched (e.g., a correct response with incorrect reason, or vice-versa), or "I+I" (0 pts) if both the response and reason were incorrect. Table 3 provides samples of students' responses scored as "C+C", "MM" and "I+I', as well the composite "reasoning score", "RS", which is the weighted average of the responses. For question effectiveness, the responses are tabulated discretely on a scale of 1 to 5 (see Section 2), from which the researchers calculated the weighted averages of these responses.

The preliminary data to date are provided in Table 4. Data are separated by problem, institution, gender, and category. Due to limitations of the study populations, the only demographic disaggregation presented here is based on binary gender identification. No statistical analysis has yet been performed, but selected items in which female students have over- or under-performed male students are highlighted in green or red, respectively.

Table 3. Samples of students' responses to question 5134

Response	Reason	Score
FBD of member DBE	By taking a sum of moments at point D on member DBE you can ignore the forces from the pin connection at D and solve for the reaction force at B. With this FBD your force acting on E will be given as F in the problem which leaves you with just one unknown.	C+C
FBD of member DBE	You can figure out the moment about B first then use the summation of forces in the X direction which will give us the force B in terms of F.	MM
FBD of member ABC including pulley C	The force at B can be solved for using an FBD of ABC because BD is a two-force member. Then you can do sum of Moments about A to determine the magnitude of the BD force.	I+I

Table 4. Summary of Results by Problem, Institution, Gender, and Category.

Table 4a.	**Correctness & Confidence**				**Reasoning**					**Effectiveness**		
ID 4550	**NR**	**NC**	**%C**	**Conf**	**NR**	**C+C**	**MM**	**I+I**	**RS**	**NR**	**UQ**	**TD**
UPR Mayagüez	81	27	33.3%	3.19	58	7	19	32	0.57	39	4.21	4.44
Female	29	9	31.0%	2.83	16	1	7	8	0.56	10	4.20	4.80
Male	52	18	34.6%	3.38	42	6	12	24	0.57	29	4.21	4.31
Whatcom (Male)	14	9	64.3%	3.11	14	5	4	5	1.00	0	n/a	n/a
Elizabethtown	35	13	37.1%	3.60	35	1	10	24	0.34	27	4.22	3.93
Female	5	2	40.0%	3.20	5	1	1	3	0.60	5	3.80	3.60
Male	30	11	36.7%	3.67	30	0	9	21	0.30	22	4.32	4.00
Oklahoma State	96	35	36.5%	3.45	87	15	21	51	0.59	69	3.67	3.37
Female	22	10	45.5%	3.17	19	4	8	7	0.84	14	3.86	3.07
Male	74	25	33.8%	3.53	68	11	13	44	0.51	53	3.62	3.45

NR = Number responding per category; **NC** and **%C** = Number and percentage of correct responses; **Conf** = average confidence (5 = max, 1 = min); **C+C** = correct response and correct reasoning; **MM** = mismatch between correctness of response and correct reasoning; **I+I** = incorrect response and incorrect reasoning; **RS** = composite reasoning score (Max = 2, Min = 0); **UQ** = average score for "I **Understood** what this **Question** was asking." (5 = max, 1 = min); **TD** = average score for "Trying to answer this question made me **Think Deeply** about course material." (5 = max, 1 = min).

Table 4b.	**Correctness & Confidence**				**Reasoning**					**Effectiveness**		
ID 4606/4758	**NR**	**NC**	**%C**	**Conf**	**NR**	**C+C**	**MM**	**I+I**	**RS**	**NR**	**UQ**	**TD**
UPR Mayagüez	65	23	35.4%	2.92	33	4	6	23	0.42	24	4.17	4.67
Female	18	9	50.0%	2.72	6	1	1	4	0.50	5	4.40	4.80
Male	47	14	29.8%	3.00	27	3	5	19	0.41	19	4.11	4.63
Whatcom (Male)	5	1	20.0%	3.80	5	0	1	4	0.20	1	4.00	4.00
Elizabethtown	35	12	34.3%	3.37	35	4	9	22	0.49	27	4.70	4.30
Female	5	3	60.0%	3.60	5	1	2	2	0.80	4	4.75	4.00
Male	30	9	30.0%	3.33	30	3	7	20	0.43	23	4.70	4.35
Oklahoma State	97	41	42.3%	3.16	85	18	16	51	0.61	69	4.36	3.83
Female	22	7	31.8%	2.83	19	4	2	13	0.53	15	4.53	4.00
Male	75	34	45.3%	3.26	66	14	14	38	0.64	54	4.31	3.78

NR = Number responding per category; **NC** and **%C** = Number and percentage of correct responses; **Conf** = average confidence (5 = max, 1 = min); **C+C** = correct response and correct reasoning; **MM** = mismatch between correctness of response and correct reasoning; **I+I** = incorrect response and incorrect reasoning; **RS** = composite reasoning score (Max = 2, Min = 0); **UQ** = average score for "I **Understood** what this **Question** was asking." (5 = max, 1 = min); **TD** = average score for "Trying to answer this question made me **Think Deeply** about course material." (5 = max, 1 = min).

Table 4c.	Correctness & Confidence				Reasoning					Effectiveness		
ID 5134	**NR**	**NC**	**%C**	**Conf**	**NR**	**C+C**	**MM**	**I+I**	**RS**	**NR**	**UQ**	**TD**
UPR Mayagüez	54	14	25.9%	3.20	32	2	7	23	0.34	23	4.39	4.57
Female	16	6	37.5%	2.19	6	1	2	3	0.67	5	4.80	4.80
Male	38	8	21.1%	3.62	26	1	5	20	0.27	18	4.28	4.50
Whatcom (Male)	7	2	28.6%	3.71	7	1	1	5	0.43	5	4.20	4.40
Elizabethtown	35	19	54.3%	3.54	35	17	2	16	1.03	27	4.59	4.48
Female	5	1	20.0%	3.20	5	1	0	4	0.40	5	4.60	4.40
Male	30	18	60.0%	3.60	30	16	2	12	1.13	22	4.59	4.50

NR = Number responding per category; **NC** and **%C** = Number and percentage of correct responses; **Conf** = average confidence (5 = max, 1 = min); **C+C** = correct response and correct reasoning; **MM** = mismatch between correctness of response and correct reasoning; **I+I** = incorrect response and incorrect reasoning; **RS** = composite reasoning score (Max = 2, Min = 0); **UQ** = average score for "I **Understood** what this **Question** was asking." (5 = max, 1 = min); **TD** = average score for "Trying to answer this question made me **Think Deeply** about course material." (5 = max, 1 = min).

Table 4d.	Correctness & Confidence				Reasoning					Effectiveness		
ID 4497	**NR**	**NC**	**%C**	**Conf**	**NR**	**C+C**	**MM**	**I+I**	**RS**	**NR**	**UQ**	**TD**
UPR Mayagüez	25	5	20.0%	4.00	15	2	0	13	0.27	12	4.75	4.33
Female	5	3	60.0%	3.20	1	1	0	0	2.00	1	5.00	5.00
Male	20	2	10.0%	4.20	14	1	0	13	0.14	11	4.73	4.27
Whatcom (Male)	12	8	66.7%	2.77	12	3	5	4	0.92	5	4.20	4.40
Elizabethtown	34	3	8.8%	3.59	34	2	1	31	0.15	26	4.35	3.69
Female	5	0	0.0%	2.80	5	0	0	5	0.00	5	4.20	3.00
Male	29	3	10.3%	3.72	29	2	1	26	0.17	21	4.38	3.86

NR = Number responding per category; **NC** and **%C** = Number and percentage of correct responses; **Conf** = average confidence (5 = max, 1 = min); **C+C** = correct response and correct reasoning; **MM** = mismatch between correctness of response and correct reasoning; **I+I** = incorrect response and incorrect reasoning; **RS** = composite reasoning score (Max = 2, Min = 0); **UQ** = average score for "I **Understood** what this **Question** was asking." (5 = max, 1 = min); **TD** = average score for "Trying to answer this question made me **Think Deeply** about course material." (5 = max, 1 = min).

4. Discussion

The first guiding question *How do faculty at different institutions employ concept-based instruction?* is directly answered by describing the different modalities of deployment. In the study, instructors employed concept questions in one of three modalities: as preparatory pre-class homework before a concept is presented in class, during class when a new concept is introduced, and as a "recap" after a concept has been covered in class. The multiple modalities used reflects the flexibility of concept-based inquiry and the CW in particular. What needs to be further explored is to understand the reasons why instructors choose different modalities. In some cases, the pre-class homework modality may be chosen to save class time needed to answer the questions and thereby leverage limited class discussion time for discussing results. On the other hand, faculty may choose to deploy questions in class to increase student engaging. A further question to explore is whether the CW questions are best used as preliminary questions as new concepts are being introduced to help students grasp a concept, or after a concept has been covered in class to help student solidify their understanding of a concept.

Regarding the second question, *How do students in different contexts respond to concept questions?*, no expectation of results or their convergence was hypothesized. Rather, the authors are in the process of studying data for any emergent patterns.

One clear finding that did emerge as uniform across all of the institutions is that female students nearly always expressed lower degrees of confidence (Conf) with their answers, regardless of whether they outperformed male students on "correctness" (%C) and/or "reasoning score" (RS) – and this occurred frequently. Perhaps this result is not surprising, as it reflects what is reported in the literature [3]–[5]. It is interesting to ponder that despite the various modalities of deployment, all instructors used the CW in what would be considered a "low stakes" scenario. The authors are not aware of any study that has attempted to disaggregate female vs. male confidence on the basis of a high stakes vs. low stakes environment. Likewise, the authors do not yet know of any study where female vs. male confidence has been evaluated specifically in the context of conceptually oriented problems.

Another clear pattern that emerges is that students across all demographics appear to have difficulty explaining a reason for the answer, even when they provide the correct response. Indeed, the composite reasoning score typically falls well below 1 (with maximum value 2). A richer analysis is in progress to perform a detailed coding and examination of the written responses, but an initial review indicates that the written responses reveal characteristics of novices, in that there is an appeal to surface features and formulae that might be out of context.

For example, in the ConcepTest 4550, some students who correctly answer that the FBD is not possible comment on things like "the forces are not labeled" or "there are four forces but only three equations". As suggested in Section 2, comments such as these, while missing the point of the problem and thus scored as "incorrect", nevertheless contain some positive reasoning – "phenomenological primitives" – that can be acknowledged to help students reframe, recontextualize, or otherwise reconsider their ideas.

ConcepTest 4497 addresses a misconception about friction that is anecdotally very well known, which is that "the harder you press, the more friction you have", and is a cousin to the simple formula "$F = \mu N$". Perhaps it is therefore not a surprise that such a problem with a "clearly obvious answer" led to a low rate of correct responses in two of the three institutions where the problem was deployed (UPR Mayagüez and Elizabethtown), and even was associated with the highest level of confidence at one institution (UPR Mayagüez). This also aligns with the "novice" approach to seek a quick formula, but it is important to pursue a deeper analysis to understand if the root of this issue is a fundamental understanding of the concept of friction, or rather, an unwillingness to apply the prior, fundamental concept of equilibrium.

There appear to be no direct correlations between question correctness, confidence, or effectiveness, neither within an institution nor across institutions. However, it might be said that in some cases, particularly at UPR Mayagüez, the expression of "TD" is optimistic compared to the baseline correctness. It is not clear whether this is a genuine impression or if it is biased in favor of "telling the instructor what they want to hear".

The authors also note that the apparently low performance rates (%C and RS) are not an indictment against concept-based instruction or the use of written explanations. It must be acknowledged that such instructional techniques are not yet normalized across engineering education. The employment of these techniques, when coupled with ample instructor-student dialogue, plausibly helps students to develop their critical thinking skills.

5. Conclusions and Future Work

This work in progress reflects a data analysis in broader study that is designed to understand contextual factors that influence both faculty and student attitudes and experiences with concept-based instruction using the Concept Warehouse. Based on data analysis of student responses to questions in the CW, four preliminary conclusions may be drawn from this work:

- Instructors use concept questions in three different modes: as preparatory pre-class homework before a concept is presented in class, during class when a new concept is introduced, and as a "recap" after a concept has been covered in class.
- Female students express lower confidence in their answers, compared with male students, regardless of whether or not the out-perform the male students, and regardless of institutional setting;
- Students have difficulty providing essential and "correct" explanations for phenomena, even when they provide the correct response, regardless of institutional setting or gender identification.
- There appear to be no correlations between student performance, confidence, and perception of question effectiveness.

A variety of direct activities can be undertaken to sharpen the results of this work. Concerning the modality question, instructors' motivations for using a particular modality need to be examined. Also, the effectiveness of using concept questions before or during introduction of a new concept versus after student have experience applying a concept should be examined.

Concerning student performance on concept questions, a formal statistical analysis should be conducted once additional data is collected. This analysis is not expected reveal any stable correlations between performance, confidence, and perception; but it will identify statistically significant variations in student performance. As more data is collected it may be possible to look at variations based on ethnicity and race as well as gender expression.

Another avenue is to undertake a more detailed analysis of students' written explanations to identify patterns, either germane to a particular problem, or to more broadly understand the general situation that occurs when students attempt to write explanations. Along with this, an improved rating for reasoning that will be more nuanced, and which will incorporate ratings from multiple researchers, is being considered.

Finally, as noted, another phase of this research is underway to explore faculty and student experiences and attitudes more directly, via survey and interviews. It is expected that this work will reveal more insights regarding usage patterns, levels of engagement, and usefulness that will inform the practice of concept-based instruction.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. Grant Nos. 1821638, 1821439, and 1821445. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] M. Koretsky, S. Nolen, B. Self, C. Papadopoulos, J. Widmann, M. Prince, D. Dal Bello, "For Systematic Development of Conceptests for Active Learning," in *Proceedings EDULEARN 19 (11th Annual International Conference on Education and New Learning Technologies)*, 2019, vol. 1, pp. 8882–8892.

[2] M. D. Koretsky, B. J. Brooks, R. M. White, and A. S. Bowen, "Querying the Questions: Student Responses and Reasoning in an Active Learning Class," *J. Eng. Educ.*, vol. 105, no. 2, pp. 219–244, Apr. 2016.

[3] D. Chachra and D. Kilgore, "Exploring gender and self-confidence in engineering students: A multi-method approach," *ASEE Annu. Conf. Expo. Conf. Proc.*, 2009.

[4] M. Besterfield-Sacre, M. Moreno, L. J. Shuman, and C. J. Atman, "Gender and ethnicity differences in freshmen engineering student attitudes: A cross-institutional study," *J. Eng. Educ.*, vol. 90, no. 4, pp. 477–489, 2001.

[5] C. L. Baird and J. R. Keene, "Closing the Gender Gap in Math Confidence : Gender and Race / Ethnic Similarities and Differences in the Effects of Academic Achievements among High Math Achievers," *Int. J. Gender,Science Technol. .*, vol. 10, no. 3, pp. 378–410, 2018.