2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
Paper ID #40193

Context Matters: Continued Study of Results of Common Concept Questions at Several Diverse Institutions

Dr. Christopher Papadopoulos, University of Puerto Rico, Mayaguez

Christopher Papadopoulos is Professor in the Department of Engineering Sciences and Materials at the University of Puerto Rico, Mayaguez Campus (UPRM).

Prof. Eric Davishahl, Whatcom Community College

Eric Davishahl serves as professor and engineering program coordinator at Whatcom Community College in northwest Washington state. His teaching and research interests include developing, implementing and assessing active learning instructional strategies and auto-graded online homework. Eric has been an active member of ASEE since 2001. He was the recipient of the 2008 Pacific Northwest Section Outstanding Teaching Award and currently serves on the ASEE Board of Directors as Zone IV Chair.

Dr. Jean Carlos Batista Abreu, Elizabethtown College

Jean Batista Abreu earned his Ph.D. and M.S.E. at the Johns Hopkins University, M.S. at the University of Puerto Rico, and B.S.E. with Honors at the Pontificia Universidad CatÃ³lica Madre y Maestra, all in Civil Engineering. Prior to joining Elizabethtown

Dr. Brian P. Self, California Polytechnic State University, San Luis Obispo

Brian Self obtained his B.S. and M.S. degrees in Engineering Mechanics from Virginia Tech, and his Ph.D. in Bioengineering from the University of Utah. He worked in the Air Force Research Laboratories before teaching at the U.S. Air Force Academy for seven years. Brian has taught in the Mechanical Engineering Department at Cal Poly, San Luis Obispo since 2006. During the 2011-2012 academic year he participated in a professor exchange, teaching at the Munich University of Applied Sciences. His engineering education interests include collaborating on the Dynamics Concept Inventory, developing model-eliciting activities in mechanical engineering courses, inquiry-based learning in mechanics, and design projects to help promote adapted physical activities. Other professional interests include aviation physiology and biomechanics.

Prof. Dominic J. Dal Bello, Allan Hancock College

Dom Dal Bello is Professor of Engineering at Allan Hancock College (AHC), a California community college between UC Santa Barbara and Cal Poly San Luis Obispo. At AHC, he is Department Chair of Mathematical Sciences, Faculty Advisor of MESA (the Mathematics, Engineering, Science Achievement Program), and Principal/Co-Principal Investigator of several National Science Foundation projects (S-STEM, LSAMP, IUSE). In ASEE, he is chair of the Two-Year College Division, and Vice-Chair/Community Colleges of the Pacific Southwest Section. He received the Outstanding Teaching Award for the ASEE/PSW Section in 2022.

Dr. Kurt M. Degoede, Elizabethtown College

Professor of Engineering and Physics, Elizabethtown College. His research interests in biomechanics include developing clinical instruments for rehabilitation. Dr. DeGoede teaches upper-level undergraduate mechanical engineering and design courses and the

Dr. Anna K. T. Howard, North Carolina State University at Raleigh

Anna Howard is a Teaching Professor at NC State University in Mechanical and Aerospace Engineering where she has led the course redesign effort for Engineering Statics. She received her Ph.D. from the Rotorcraft Center of Excellence at Penn State University.

Dr. Azize Akcayoglu
Dr. Hadas Ritz, Cornell University

Hadas Ritz is a senior lecturer in Mechanical and Aerospace Engineering, and a Faculty Teaching Fellow at the James McCormick Family Teaching Excellence Institute (MTEI) at Cornell University, where she received her PhD in Mechanical Engineering in 2008. Among other teaching awards, she received the 2021 ASEE National Outstanding Teaching Award.

Dr. William A. Kitch, Angelo State University

Dr. Kitch is Professor and Chair of the David L. Hirschfeld Department of Engineering at Angelo State University. Before starting his academic career he spent 24 years as a practicing engineer in both the public and private sector. He is a registered pr

Context Matters: Continued Study of Results of Common Concept Questions at Several Diverse Institutions

1. Introduction

Concept-based instruction is an approach to deploy "concept questions" which are qualitative and designed to elicit patterns of thought that complement or reinforce those required for procedural questions. Typically, concept questions are multiple choice with one "correct" answer among several "attractive distractors". However, some concept questions may, by design, have "multiple defensible responses", to engender debate and deeper discussion about multiple solution pathways, underlying assumptions, or other contextual details. Also, the use of concept questions is arguably most effective when written explanations of answers are also collected, so as to better understand students' reasoning, including the possibility that an "incorrect" answer reveals some measure of conceptual understanding (sometimes referred to as a "phenomenological primitive"). Finally, use of concept questions is part of an evolutionary process of faculty development, in which the deployment, review of explanations, and feedback, is an ongoing process oriented toward effective teaching and learning outcomes (Koretsky et al., 2019).

A Community of Practice (CoP) of mechanics instructors from several diverse institutions (ranging in size, demographics, and identity), has been formed to use the Concept Warehouse (CW) as a platform to create, deploy, and assess the results of concept questions in Statics and Dynamics. The CW is an online tool that contains several thousand concept questions, called "ConcepTests", that range over several topics in engineering, including approximately 800 in mechanics. The CW allows the instructor to deploy the ConcepTests in a variety of modalities, including online or offline, in-class or out of class, and with response time allocated to be "immediate" (say 2-5 minutes during class) or "extended" (say several hours or days as a preparatory or exploratory exercise).

The CoP has two teams, one for Statics, and one for Dynamics. During the Fall 2022 semester, each Statics or Dynamics team member assigned the same four "common questions" from the CW, at the point and in the modality appropriate to their course. The following data was collected: the answer to the question, corresponding written explanations (i.e., to explain or justify the chosen answer), and immediate feedback (e.g., confidence and impressions as to the usefulness of the question). A small portion of students also participated in follow-up interviews.

This work is the sequel to a work-in-progress (WIP) article published and presented at the 2022 Annual Conference & Exposition (Papadopoulos et al., 2022), conducted by four faculty teaching Statics. In this study, use of the same four common Statics questions from the WIP is repeated, while four Dynamics questions are added (one of these is identical to one of the Statics questions). The WIP reported two general findings:

- Across all institutions, and independently of correctness of their answers, female students consistently reported lower confidence in their answers.

- Among students selecting correct responses, only about one third to one half expressed reasoning that was considered “correct”. Nevertheless, many “incorrect” answers contained portions of reasoning that suggested that some core ideas were being expressed, allowing for the possibility of further discussion to build understanding.

This study will inquire as to whether these trends persist. In addition, the group of authors has matured and expanded, and through a regular meeting Community of Practice, they have debated details of question phrasing to larger questions of how to make use of student responses. Additional issues that are addressed in this article relate to the effect of timing, repetition, and modality of deployment on student performance.

2. Institutional Profiles

Tables 1 and 2 provide descriptions of the participating institutions and the modality of deployment of the CW questions.

Table 1. Summary of Institutions and Modalities for Statics.

Institution	Description
S1: University of Puerto Rico, Mayagüez (UPRM)	Public, mid-sized, urban, bilingual, HSI. Primarily deployed in class after substantial discussion on topic. Did not consistently redeploy, so initial results are given.
S2: Whatcom Community College	Public, mid-sized, suburban community college engineering transfer program. CTs 7059, 4756 and 4497 deployed in class using peer instruction. CT 5134 deployed as homework pre and post related content coverage through class activities and homework.
S3: Elizabethtown College (E-town)	Private, small, rural, liberal arts. Deployed at the start of class session. CTs 4606 and 5134 were asked after the topic was discussed. CT 4497 was asked before the topic was introduced.
S4: North Carolina State University	Public, large R1. Questions asked in weekly quizzes and the final exam (summative assessments.) Deployed outside CW platform with identical questions.
S5: Allan Hancock College	Public, mid-sized, rural, HSI, community college transfer program. Questions deployed as homework, with questions discussed as a class at the beginning of the next class session.
S6: Angelo State University	Public, mid-sized, rural, HSI, four-year engineering program. Questions deployed as pre-class concept questions to facilitate in-class discussions.

Table 2. Summary of Institutions and Modalities for Dynamics.

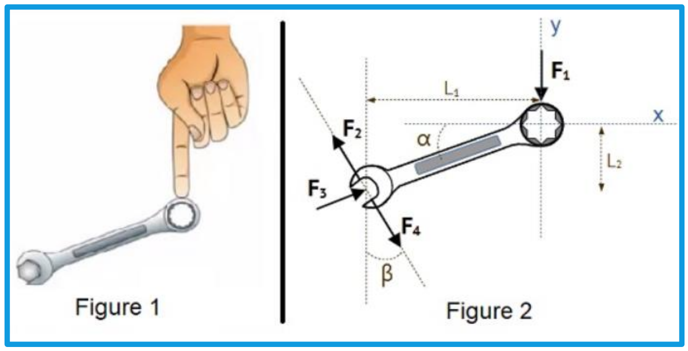
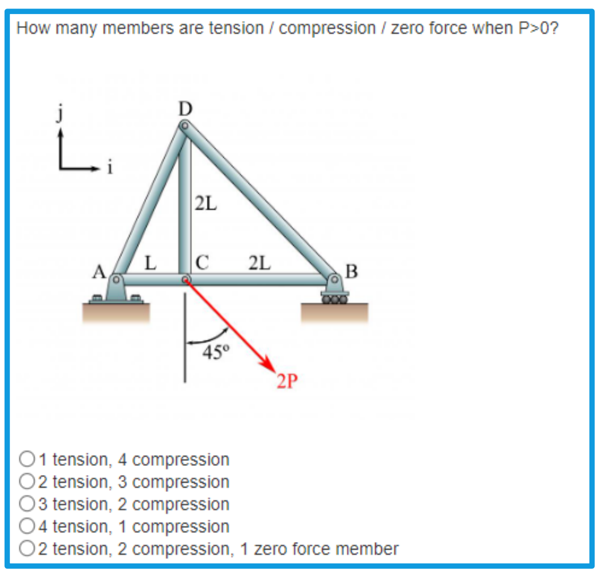
Institution	Description
D1: Cal Poly San Luis Obispo (Cal Poly SLO)	Public, mid-sized, rural polytechnic. Deployed in class.
D2: Elizabethtown College (E-town)	Private, small, rural. Deployed at start of class session before topic was discussed. Foundation for in-class discussion, but questions were not redeployed in CW. Not all students provided reasoning.

D3: Allan Hancock College	Public, mid-sized, rural, HSI, community college transfer program. Questions deployed as homework, with questions discussed as a class at the beginning of the next class session.
D4: Angelo State University	Public, mid-sized, rural, HSI, four-year engineering program. Questions deployed as homework towards the end of the semester. Data from this cohort is from Spring 2023.

3. Description of Common Questions

Tables 3 and 4 summarize the four common statics questions and the four common dynamics questions, respectively. Note that one question, ID 4497, is common to both groups.

Table 3. The Four Common Statics Questions.

ID	Topic and Text	ConcepTest
7059	**Wrench** A force is applied to a wrench that grips a hex-head bolt, as shown in Figure 1. A proposed free body diagram is shown in the Figure 2. Is the free body diagram suitable for analyzing this problem? ● Yes ● No ● Cannot be determined from the given information	 *The intention is for students to notice that forces F2, F3, and F4 intersect at a common point, leading them to consider the moment equilibrium of the entire wrench.*
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 ● 4 Tension, 1 Compression ● 2 Tension, 2 Compression, 1 Zero-force member	 *The intention is for students to draw mental or actual FBDs of various joints, make qualitative determinations about the modality of the member, and then continue to another joint to complete the analysis.*

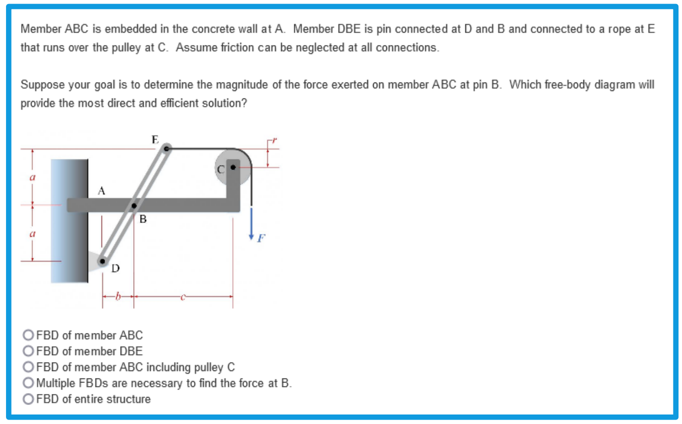
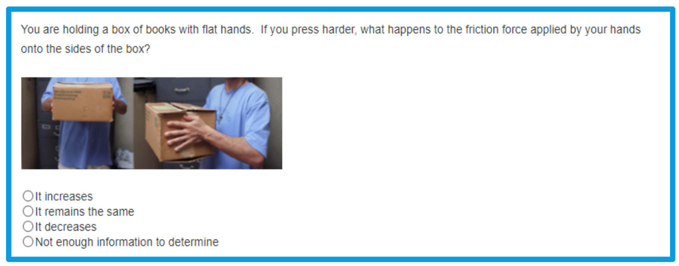
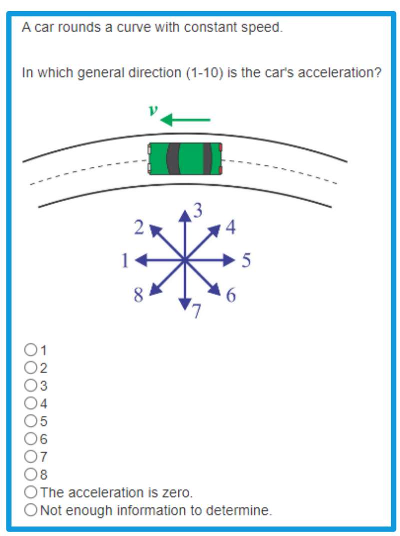
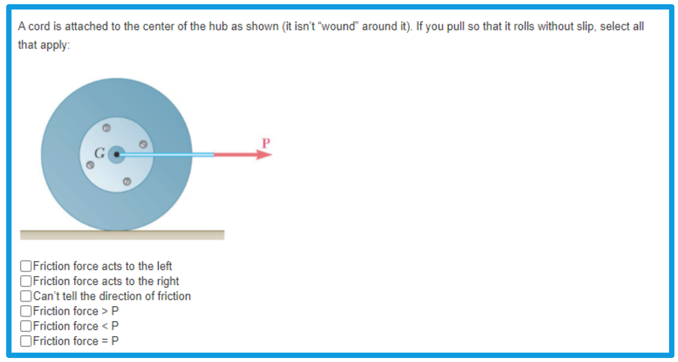
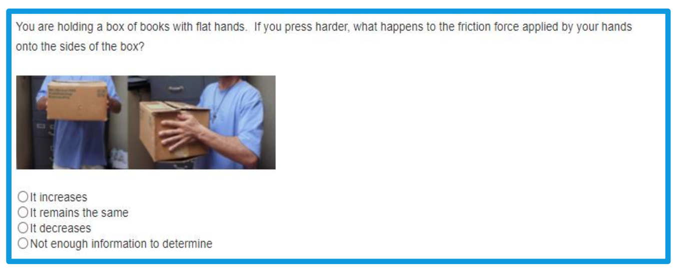
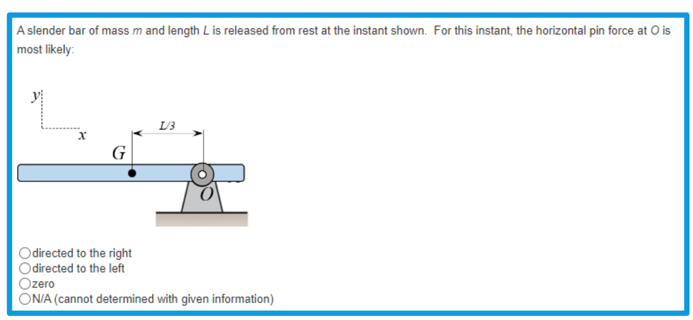
5134	**Frames and Machines** "Member ABC is embedded in the concrete wall at A. Member DBE is pin connected at D and B is connected to a rope at E that runs over the pulley at C. Assume that friction can be neglected at all connections. Suppose your goal is to determine the magnitude of the force exerted 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	**Box with Friction** "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 to confront a simple situation in which the common law "F = μN" does not apply, and to realize the importance of drawing a simple FBD and applying equilibrium.*

Table 4. The Four Common Dynamics Questions

7077, 7078, 7079 Formerly 5844, 5845, 5846	**Car on a Curve 1 (7077)** "A car rounds a curve with constant speed. In which general direction (1-8) is the car's acceleration?" • 1 • 2 • 3 • 4 • 5 • 6 • 7 • 8 • The acceleration is zero • Not enough information **Car on a Curve 2 (7078)**	

	"A car rounds a curve with speed v, and is slowing down. In which general direction (1-8) is the car's acceleration?" **Car on a Curve 3 (7079)** "A car rounds a curve with speed v, and is speeding up. In which general direction (1-8) is the car's acceleration?"	*The intention is for students to recognize that (1) there is always a normal acceleration when moving on a curved path, and there may or may not be a tangential acceleration.* *Note: Effective Spring 2023, Concept Tests 7077, 7078, and 7079 replaced 5844, 5845 and 5846 from the WIP, respectively. Options 9 and 10 were added did not appear in the original versions: (9) The acceleration is zero. (10) Not enough information to determine.*
4618	**Spool Center** "A cord is attached to the center of the hub as shown (it isn't "wound" around it). If you pull so that it rolls without slip, select all that apply:" • Friction force acts to the left • Friction force acts to the right • Can't tell the direction of friction • Friction force > P • Friction force < P • Friction force = P	 *The intention is for students to understand the relationships between forces and linear accelerations, moments and angular accelerations, and between linear and angular accelerations.*
4497	**Box with Friction** "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 to confront a simple situation in which the common law "F = µN" does not apply, and to realize the importance of drawing a simple FBD and applying equilibrium.*
4711	**Slender Bar** "A slender bar of mass m and length L is released from rest at the instant shown. For this instant, the horizontal pin force at O is most likely:" • directed to the right • directed to the left • zero • N/A (cannot determined with given information)	 *The intention is for students to understand that, for this case, if the angular velocity is zero (at rest), then the centripetal acceleration is zero, and thus no (horizontal) force on the rod is required.*

In addition to the questions themselves, the Concept Warehouse is designed to solicit additional responses, including written explanations of student reasoning, and ratings of confidence and question effectiveness. Table 5 provides the format for these questions.

Table 5. Collection of Student Explanations and Ratings for Confidence and Question Effectiveness.

Student Explanations and Confidence	Question Clarity and Effectiveness
Please explain your answer in the box below. Please rate how confident you are with your answer. substantially unsure ○ moderately unsure ○ neutral ○ moderately confident ○ substantially confident ○ *Note: The CW allows the instructor to read student explanations either in real time as responses are being submitted or after the question is closed.*	Please help us assess the effectiveness of this question by answering the items below: I understood what this question was asking. strongly disagree ○ moderately disagree ○ neutral ○ moderately agree ○ strongly agree ○ Explain your response to the item above. Trying to answer this question made me think deeply about course material. strongly disagree ○ moderately disagree ○ neutral ○ moderately agree ○ strongly agree ○ Explain your response to the item above.

4. General Results

For this study, we asked students to complete all questions described in Table 5, to better understand the context of their responses. From this data, we substantiate the results from the previously cited WIP (Papadopoulos et al., 2022), male vs. female confidence, and correctness of response vs. correctness of corresponding reasoning. That is, male students generally report higher confidence, and many students who select a correct answer are unable to provide correct or sufficient reasoning to justify the answer.

4a. Confidence as a Function of Gender. As can be seen from Table 6, in all cohorts, with only minor exceptions, male students nearly always report higher confidence in their answers than female students, regardless of whether their actual performance was higher or not. Students who did not identify as female or male were excluded because their small numbers made it difficult to maintain their anonymity. Also note that cohorts S2 (Whatcom Community College) and D3 (Allan Hancock) consist only of male students.

Table 6. Confidence vs. Gender

Question >	7059 Wrench		4756 Truss		5134 Frame		4497 Box with Friction	
Institution	Correct	Conf	Correct	Conf	Correct	Conf	Correct	Conf
S1 UPRM	12/27 (44%)	n/a	3/6 (50%)	2.22	10/27 (37%)	3.14	2/27 (7%)	2.85
Male	9/20 (45%)	n/a	2/6 (33%)	2.33	6/19 (32%)	3.41	2/20 (10%)	2.95
Female	3/7 (43%)	n/a	1/3 (33%)	2.00	4/8 (50%)	2.50	0/7 (1%)	2.57
S2 Whatcom								
Male	7/12 (58%)	n/a	10/14 (71%)	4.00	5/14 (36%)	4.07	10/15 (67%)	4.13
S3 E-town	n/a	n/a	13/24 (54%)	3.19	15/25 (60%)	4.16	9/25 (36%)	4.00
Male	n/a	n/a	10/18 (56%)	3.47	14/19 (74%)	4.16	5/19 (26%)	4.21
Female	n/a	n/a	3/6 (50%)	2.33	1/6 (17%)	4.17	4/6 (67%)	3.33
S5 Allan Hancock	7/9 (78%)	3.71	4/11 (36%)	4.00	1/13 (8%)	3.68	3/11 (27%)	4.27
Male	6/8 (75%)	3.80	3/9 (33%)	4.22	1/11 (9%)	3.71	3/10 (30%)	4.20
Female	1/1 (100%)	3.00	1/2 (50%)	3.00	0/2 (0%)	3.50	0/1 (0%)	5.00
S6: Angelo State	8/19 (42%)	2.58	8/14 (57%)	2.91	5/16 (31%)	2.82	5/12 (42%)	3.42
Male	8/18 (44%)	2.56	8/13 (62%)	3.13	4/13 (31%)	2.93	5/11 (45%)	3.73
Female	0/1 (0%)	3.00	0/1 (0%)	0.00	1/3 (33%)	2.33	0/1 (0%)	0.00
	5844 Car on Curve 1		5618 Spool Center		4711 Slender Bar		4497 Box with Friction	
D1 Cal Poly SLO	17/18 (94%)	4.43	n/a	n/a	9/19 (47%)	3.17	14/17 (82%)	4.59
Male	9/9 (100%)	4.57	n/a	n/a	6/10 (60%)	3.50	6/9 (67%)	5.00
Female	8/9 (89%)	4.29	n/a	n/a	3/9 (33%)	2.80	8/8 (100%)	4.13
D2 E-town	7/28 (25%)	2.68	20/35 (57%)	2.96	6/31 (19%)	2.70	22/35 (63%)	3.20
Male	6/21 (29%)	2.80	15/28 (54%)	3.10	3/24 (13%)	2.70	19/29 (66%)	3.20
Female	1/7 (14%)	2.30	5/7 (71%)	2.40	3/7 (43%)	2.70	3/6 (50%)	3.20
D3 Allan Hancock								
Male	n/a	n/a	2/6 (33%)	2.50	5/6 (83%)	3.67	2/10 (20%)	3.90
D4 Angelo State	2/14 (14%)	4.60	7/14 (50%)	4.86	9/14 (64%)	4.39	n/a	n/a
Male	1/13 (8/%)	4.72	6/13 (46%)	5.00	8/13 (61%)	4.50	n/a	n/a
Female	1/1 (100%)	3.00	1/1 (100%)	3.00	1/1 (100%)	3.00	n/a	n/a

Notes. Correct = number of correct responses/number of total responses; raw data is provided in fractional form, and the corresponding percentage appears in (parentheses). Conf = average confidence score of cohort, with 5 = substantially confident, 4 = moderately confident, 3 = neutral, 2 = moderately unconfident, 1 = substantially unconfident.

4b. Relation Between Correct Response and Correct Explanation. For each respondent, the correctness of the answer[1] was compared with the quality of the corresponding written response

[1] By design, ConcepTests might have multiple defensible responses, with the objective to promote debate and inquiry. Therefore the notion of the existence of single correct answer can be unhelpful. However in the set of questions examined in this study, the questions have a single best answer that is designated as "correct".

to determine if the student adequately justified the answer. Based on manual examination of written responses, the authors judged whether the reasoning provided correctly justified a correct answer using a binary scale (either yes or no). The authors held meetings to calibrate how they would judge the correctness of the response in a uniform manner. Table 7 provides the results comparing correctness of response to the provided justification. As before, cohorts S2 (Whatcom Community College) and D3 (Allan Hancock) consist only of male students; the cohort S4 (North Carolina State) has male and female students aggregated.

Table 7. Correctness of Response vs. Correctness of Explanation

Qu >	7059 Wrench		4756 Truss		5134 Frame		4497 Box with Friction	
Inst	CA	CR	CA	CR	CA	CR	CA	CR
S1 Male Female	12/27 (44%) 9/20 (45%) 3/7 (43%)	7/12 (58%) 5/9 (56%) 2/3 (67%)	3/6 (50%) 2/3 (33%) 1/3 (33%)	n/a n/a n/a	10/27 (37%) 6/19 (32%) 4/8 (50%)	9/10 (90%) 5/6 (83%) 4/4 (100%)	2/27 (7%) 2/20 (10%) 0/7 (0%)	0/2 (0%) 0/2 (0%) 0/0 (--)
S2 Male	7/12 (58%)	1/7 (14%)	10/14 (71%)	9/10 (90%)	5/14 (36%)	5/5 (100%)	10/15 (67%)	10/10 (100%)
S3 Male Female	n/a n/a n/a	n/a n/a n/a	13/24 (54%) 10/18 (56%) 3/6 (50%)	7/13 (54%) 5/10 (50%) 2/3 (67%)	15/25 (60%) 14/19 (74%) 1/6 (17%)	12/15 (80%) 12/14 (86%) 0/1 (0%)	9/25 (36%) 5/19 (26%) 4/6 (67%)	7/9 (78%) 3/5 (60%) 4/4 (100%)
S4 Male+ Female	267/327 (82%)	148/267 (55%)	251/305 (82%)	n/a	129/315 (41%)	n/a	213/310 (69%)	195/213 (92%)
S5 Male Female	7/9 (78%) 6/8 (75%) 1/1 (100%)	5/7 (71%) 4/6 (67%) 1/1 (100%)	4/11 (36%) 3/9 (33%) 1/2 (50%)	2/4 (50%) 2/3 (67%) 0/1 (0%)	1/13 (8%) 1/11 (9%) 0/2 (0%)	0/1 (0%) 0/1 (0%) 0/0 (--)	3/11 (27%) 3/10 (30%) 0/1 (0%)	0/3 (0%) 0/3 (0%) 0/0 (--)
S6 Male Female	8/19 (42%) 8/18 (44%) 0/1 (0%)	3/8 (38%) 3/8 (38%) 0/0 (--)	8/14 (57%) 8/13 (62%) 0/1 (0%)	5/8 (63%) 5/8 (63%) 0/0 (--)	5/16 (31%) 4/13 (31%) 1/3 (33%)	4/5 (80%) 3/4 (75%) 1/1 (100%)	5/12 (42%) 5/11 (45%) 0/1 (0%)	2/5 (40%) 2/5 (40%) 0/0 (--)
	5844 Car on Curve 1		5618 Spool Center		4711 Slender Bar		4497 Box with Friction	
D1 Male Female	17/18 (94%) 9/9 (100%) 8/9 (78%)	15/17 (88%) 9/9 (100%) 6/8 (75%)	n/a n/a n/a	n/a n/a n/a	9/19 (47%) 6/10 (60%) 3/9 (33%)	7/12 (58%) 5/6 (83%) 2/6 (33%)	14/17 (82%) 6/9 (67%) 8/8 (100%)	11/14 (79%) 3/6 (50%) 8/8 (100%)
D2 Male Female	7/28 (25%) 6/21 (29%) 1/7 (14%)	4/7 (57%) 4/6 (67%) 0/1 (0%)	20/35 (57%) 15/28 (54%) 5/7 (71%)	9/20 (45%) 8/15 (53%) 1/5 (20%)	6/31 (19%) 3/24 (13%) 3/7 (43%)	5/6 (83%) 2/3 (67%) 3/3 (100%)	22/35 (63%) 19/29 (66%) 3/6 (50%)	8/22 (36%) 6/19 (32%) 2/3 (67%)
D3 Male	n/a	n/a	2/6 (33%)	0/2 (0%)	5/6 (83%)	0/5 (0%)	2/10 (20%)	0/2 (0%)
D4 Male Female	2/14 (14%) 1/13 (8%) 1/1 (100%)	n/a n/a n/a	7/14 (50%) 6/13 (33%) 1/1 (100%)	n/a n/a n/a	9/14 (64%) 8/13 (61%) 1/1 (100%)	n/a n/a n/a	n/a n/a n/a	n/a n/a n/a
Notes. Correct Answer CA = number of correct responses/number of total responses. Correct Reasoning CR = number of adequate justifications/number of correct responses. Raw data is provided in fractional form; corresponding percentage appears in (parentheses).								

The results can vary across question, institution, and other demographics, but in general, typically not more than two thirds of students who select the correct answer can adequately justify it. A similar phenomenon was observed in (Koretsky et al., 2016). This suggests that caution must be used to interpret correct responses to multiple choice concept questions as representing sufficient understanding of the concept.

In the case of instructor S4 (author Howard), the correlation coefficient between the correct answers and the correct reasoning varied significantly across problems: the Box with Friction had a very high correlation (R = 0.81) between students getting the answer correct and the reasoning correct, while the Wrench problem had a much larger variety of reasons that the students found the FBD given insufficient. These two questions where reasoning was required had correct response rates of 82% and 41%, respectively, while the other two questions where reasoning was not required, the Truss and Frame, had correct response rates of 82% and 69%, respectively. Therefore, no clear effect of requiring vs. not requiring a rationale was observed. For comparison, (Koretsky et al., 2016) conclude that the presence of written responses generally correlates with higher performance.

Another possible explanation for the variation within a cohort might be that different instructors emphasize different problem-solving ideas or concepts. This became apparent through some discussions among the authors in preparing this article.

Another issue related to the collection of written responses is the 'sharpness' of the question. Students often will comment on surface features rather than more fundamental concepts; for example, in ID 7059 (Equilibrium of the Wrench), several students comment on whether the FBD has adequate dimensioning, rather than on the fundamental notion of the equilibrium of the wrench. So, some problems will less sharply elicit thought on a concept if they lend themselves more to 'surface feature responses'. This suggests that writers of questions must become skilled and steering attention to the concept at hand.

Regardless of these and other caveats, the authors maintain that even explanations that are flawed might indicate a seed of a sound idea, and discussion of these ideas can lead to a deeper investigation of the topic. Sometimes, providing written explanations first liberates some students to voice their ideas in class when they would otherwise have chosen to remain silent.

5. Other Results and Observations.

In addition to the general results reported in the previous section, and given the diversity of institutions, modalities, and other circumstances, a variety of insights and interpretations of results emerge. In this section, certain important results and observations from each participating instructor are reported.

5a. Effect of Timing of Deployment. During the previous year, and as reported in the WIP, instructor S1 (Papadopoulos) deployed the CW problems as preliminary problems to be done as part of the introduction to the course material. It was later thought that students might benefit from having some general exposure to the material before attempting the CW questions, thereby using the CW questions to revisit or reinforce ideas. Therefore, in the current data set, the CW problems were deployed in class after one or two lectures of exposure. Table 8 provides a comparison of the results for correctness and reasoning.

Table 8. Comparison of Results from S1, Fall 2021 (Pre-exposure Deployment) vs. Fall 2022 (Post-exposure Deployment)

	7059 Wrench		4756 Truss		5134 Frame		4497 Box with Friction	
	CA	CR	CA	CR	CA	CR	CA	CR
Fall 2021	27/81 (33%)	7/27 (26%)	23/65 (35%)	4/23 (17%)	14/54 (26%)	2/14 (14%)	5/25 (20%)	2/5 (40%)
Male	18/52 (35%)	6/18 (33%)	14/47 (30%)	3/14 (21%)	8/38 (21%)	1/8 (13%)	2/20 (10%)	1/2 (50%)
Female	9/29 (31%)	1/9 (11%)	9/18 (50%)	1/9 (11%)	6/16 (38%)	1/6 (17%)	3/5 (60%)	1/3 (33%)
Fall 2022	12/27 (44%)	7/12 (58%)	3/6 (50%)	n/a	10/27 (37%)	9/10 (90%)	2/27 (7%)	0/2 (0%)
Male	9/20 (45%)	5/9 (56%)	2/3 (33%)	n/a	6/19 (32%)	5/6 (83%)	2/20 (10%)	0/2 (0%)
Female	3/7 (43%)	2/3 (67%)	1/3 (33%)	n/a	4/8 (50%)	4/4 (100%)	0/7 (0%)	0/0 (--)
Notes. Correct Answer CA = number of correct responses/number of total responses. Correct Reasoning CR = number of adequate justifications/number of correct responses. Raw data is provided in fractional form; corresponding percentage appears in (parentheses).								

According to the results, students generally performed better on the ConcepTests after having had some exposure to the topic, both in terms of correctness and ability to justify the reasoning. This suggests that ConcepTests are not necessarily appropriate as part of the initial exposure to a topic, such as occurs in the Inverted Classroom format (used by the instructor), unless the level of the questions are properly calibrated.

5b. Effect of Repetition. Instructor D2 (author Davishahl) observed that most students indicated appropriate lack of confidence when recalling concepts from prerequisite course work (Physics I and Statics). Replication of the exact same ConcepTest using the Friction Box a year later indicated strong retention. This raises the following question: would that transfer to another application of the same concept?

5c. Effect of Modality. Instructor S1 (author Papadopoulos) originally developed the Box with Friction problem as part of a final exam many years ago (see Figure 1), which is substantially similar to the ConcepTest (ID 4497). In Spring 2010 and Fall 2010, the problem was required and accounted for 16/100 exam points; in Spring 2013 and Fall 2013, the problem was given as a bonus worth up to +5/100 points. This represents a mixture of high- and low-stakes conditions, but both were on paper in an open format, inviting a procedure. In contrast, both deployments for the CW, in Fall 2021 and Fall 2022, were online, though in the first case it was assigned as an out of class, preliminary exercise, with several days to complete, and in the second case, as an in-class activity, with a few minutes to complete.

Figure 1. Box with Friction Problem as Originally Deployed in a Final Exam.

To compare the results, the Final Exam questions from 2010 and 2013 were rescored to align with the discrete choices from the CW question; in particular, a "correct" designation is assigned to a student who clearly indicated that the frictional force would not change. Table 9 reports the number and rate of the correct response in each instance.

Table 9. Box with Friction Problem: Comparison of Results as Deployed in a Final Exam vs. through the CW Platform.

Deployment and Conditions	Spring 2010 Final Exam	Fall 2010 Final Exam	Spring 2013 Final Exam	Fall 2013 Final Exam	Fall 2021 CW Online Out of Class	Fall 2022 CW Online In Class
Grade Weight	**Required 16/100**	**Required 16/100**	**Bonus +5/100**	**Bonus +5/100**	**None**	**None**
Number and % correct	9/36 (25%)	16/36 (44%)	8/35 (23%)	2/21 (10%)	5/25 (20%)	2/20 (10%)
Notes: The number (and percentage) of students answering "remains the same", "does not change", or equivalent.						

According to this data, students performed better in the exam format, especially in the high-stakes case. Moreover, students who expressed the correct answer on the Final Exam (who indicated that the friction force does not change) usually included valid justification, such as with a Free Body Diagram and equilibrium equations. In contrast, after submitting their answers online in the CW, the instructor asked by show of hands if anyone drew a Free Body Diagram as part of their work; essentially no one said 'yes', and many actively voiced or shook heads to indicate 'no'.

As a further example, instructor S4 (author Howard) downloaded the four Statics questions and delivered them through Moodle, and asked all exclusively on individual-effort summative assessments. The Wrench, Box with Friction, and Truss questions were asked on weekly quizzes; the frame question appeared on the final exam. The percentage correct was higher on the weekly quizzes which were untimed. Though students were asked to turn in their scrap work, they were not required to draw free-body diagrams.

One possible explanation is that the online environment, together with the notion that concept questions are "supposed to be solved mentally" without resort to pencil & paper, conspire to discourage students from deep thinking, especially when there is an available answer that makes sense. Conversely, the written exam or quiz format, which invites an expressive answer, and which does not ostensibly frame the question as a "simple concept question", elicits deeper, more nuanced responses that tend to be more correct.

6. Conclusions

This study collected data from several instructors participating in "common questions" study to understand patterns of student performance in answering concept questions. Two findings remain relatively consistent from the prior WIP.

First, even when students answer a question correctly, i.e., they select the correct option, their written explanations are often flawed or unconvincing. This shows the limitations of using performance on ConcepTests as an accurate measure of students' conceptual understanding. But since many instructors routinely cite such performance data in published studies and internal assessments – including these authors – there is likely real value to probe more deeply and analyze open-ended responses, both to better understand what students think, and to better understand the limitations of concept question results.

The second trend that persisted is the lower confidence of female students compared to their male counterparts (confidence of students of other gender identifications was not undertaken), and this mirrors other results reported in the literature b(Baird & Keene, 2018; Besterfield-Sacre et al., 2001). Given that this phenomenon is well established, it is important to move beyond merely providing further documentation. How might instructors understand and respond to this pattern? Given that retention of women is on a par of that of men – recruitment more problematic – what might lower female confidence imply about the environment in which they learn? Is lack of confidence due to climate or students' belongingness? Increased attention is being paid to these issues, and perhaps new interventions need to be designed to increase women's self-efficacy in STEM classes.

With regard to the other observations about timing, modality, and repetition, further inquiry is necessary to establish if the results are situational or suggestive of general patterns. The authors intend to maintain working through a Community of Practice to refine the questions and generate further data to support more definitive conclusions.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. Grant Nos. 1821445, 1821638, 1821439, and 1821603. 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

Baird, C. L., & Keene, J. R. (2018). Closing the Gender Gap in Math Confidence: Gender and Race / Ethnic Similarities and Differences in the Effects of Academic Achievements among High Math Achievers. *International Journal of Gender,Science & Technology*, *10*(3), 378–410. https://genderandset.open.ac.uk/index.php/genderandset/article/view/452

Besterfield-Sacre, M., Moreno, M., Shuman, L. J., & Atman, C. J. (2001). Gender and ethnicity differences in freshmen engineering student attitudes: A cross-institutional study. *Journal of Engineering Education*, *90*(4), 477–489. https://doi.org/10.1002/j.2168-9830.2001.tb00629.x

Koretsky, M. D., Brooks, B. J., & Higgins, A. Z. (2016). Written justifications to multiple-choice concept questions during active learning in class. *International Journal of Science Education*, *38*(11), 1747–1765. https://doi.org/10.1080/09500693.2016.1214303

Koretsky, M., Nolen, S., Self, B., Papadopoulos, C., Widmann, J., Prince, M., & Dal Bello, D. (2019). For Systematic Development of Conceptests for Active Learning. *EDULEARN19 Proceedings*, *1*, 8882–8892. https://doi.org/10.21125/edulearn.2019.2205

Papadopoulos, C., Davishahl, E., Ramming, C. H., Abreu, J. C. B., & Kitch, W. A. (2022). Work in Progress: Context Matters: A Comparative Study of Results of Common Concept Questions in Statics at Several Diverse Institutions. *ASEE Annual Conference and Exposition, Conference Proceedings*. https://peer.asee.org/41627