Phrasing Matters: A Case Study in the Evolution of a Concept Question and the Effect on Student Responses

Dr. Christopher Papadopoulos, University of Puerto Rico, Mayaguez Campus

Christopher Papadopoulos is Professor of Engineering Sciences and Materials at the University of Puerto Rico, Mayagüez (UPRM). He earned B.S. degrees in Civil Engineering and in Mathematics from Carnegie Mellon University (1993) and a Ph.D. in Theoretical and Applied Mechanics at Cornell University (1999). Prior to UPRM, Papadopoulos served on the faculty in the Department of Civil Engineering and Mechanics at the University of Wisconsin, Milwaukee.

Papadopoulos has diverse interests in structural mechanics, sustainable construction materials (with emphasis in bamboo), engineering ethics, and engineering education. He is co-author of Lying by Approximation: The Truth about Finite Element Analysis, and after many years, he has finally (maybe) learned how to teach Statics, using an experiential and peer-based learning "studio" model. As part of the UPRM Sustainability Engineering initiative to develop a new bachelor's degree and curricular sequence, Papadopoulos is PI of A New Paradigm for Sustainability Engineering: A Transdisciplinary, Learner-Centered, and Diversity-Focused Approach, funded by the NSF HSI program. Papadopoulos is active in the Mechanics (former Chair), Ethics, and LEES Divisions of ASEE, and is the co-president of the UPRM Institutional Committee for General Education. He enjoys biking, swimming, cooking, and eliminating disposable plastic.

Papadopoulos endeavors to orient his research and teaching activities around advancing, peace, social equity, justice, and human wellbeing. In the words of Roberto Clemente, anytime when you have the opportunity to make a difference in the world, and you don't do it, then you are wasting your time on earth.

Dr. William A Kitch P.E., 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 professional engineer.

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 and is one of the campus leaders of Wolfpack Engineering Unleashed. She has launched and is currently chairing the College Teaching Committee for the NC State College of Engineering.

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), has served as 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. 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 Católica Madre y Maestra, all in Civil Engineering. Prior to joining Elizabethtown College, he served as a Visiting Assistant Professor of Mechanical and Civil Engineering at Bucknell University, and worked as a Structural Engineer at Acero Estrella.

Dr. Julian Ly Davis, University of Southern Indiana

Jul Davis is an Associate Professor of Engineering at the University of Southern Indiana in Evansville, Indiana. He received his PhD in 2007 from Virginia Tech in Engineering Mechanics where he studied the vestibular organs in the inner ear using finite element models and vibration analyses. After graduating, he spent a semester teaching at a local community college and then two years at University of Massachusetts (Amherst) studying the biomechanics of biting in bats and monkeys, also using finite element modeling techniques. In 2010, he started his career teaching in all areas of mechanical engineering at the University of Southern Indiana. He loves teaching all of the basic mechanics courses, and of course his Vibrations and Finite Element Analysis courses.

Prof. Eric Davishahl, Whatcom Community College

Eric Davishahl serves as professor and engineering program coordinator at Whatcom Community College in northwest Washington state. His current project involves developing and piloting an integrated multidisciplinary learning community for first-year engineering. More general teaching and research interests include designing, implementing and assessing activities for first-year engineering, engineering mechanics, and scientific computing. Eric has been an active member of ASEE since 2001. He was the recipient of the 2008 Pacific Northwest Section Outstanding Teaching Award, chaired the PNW Section 2017-2019, served on the ASEE Board of Directors as Zone IV Chair 2022-2024. and currently serves as Program Chair for the Two-Year College Division.

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 sev

Dr. Milo Koretsky, Tufts University

Milo Koretsky is the McDonnell Family Bridge Professor in the Department of Chemical and Biological Engineering and in the Department of Education at Tufts University. He received his B.S. and M.S. degrees from UC San Diego and his Ph.D. from UC Berkeley,

Phrasing Matters: A Case Study in the Evolution of a Concept Question and the Effect on Student Responses

Abstract. A total of 1,685 responses from three different versions of a ConcepTest in the Concept Warehouse are evaluated (557, 881, and 247 responses for Versions 1, 2, and 3, respectively). In all responses, students were asked to choose the correct answer and provide a written explanation of their answer choice. The underlying issue of the question is for the student to discover a lack of moment equilibrium in the provided Free Body Diagram. The primary result of the work suggests that a question phrasing with the explicit question "can the body be in static equilibrium," rather than asking if the Free Body Diagram is "possible" or "suitable," is more likely to bring student attention to the key underlying issue. Secondary effects of adding kinematic annotations to the FBD, such as axes, dimensions, and angles, were also observed.

Keywords: concept questions, written explanations, surface features, question phrasing, question wording

1. Introduction

Concept questions, also called ConcepTests, focus on promoting qualitative reasoning and sense-making, and their deployment has become a well-established means to promote active learning in engineering and other fields over the last two or three decades [1]. Development of concept questions ranges from individual instructors posing spontaneous questions based on their intuition as to what illuminates an important concept [2], to psychometrically developed studies that result in instruments known as concept inventories [3], [4]. During development of concept questions, attention is paid to presenting a clear and concise question that typically focuses on a single concept. Wording and pictorial representations play a key role in creating a "good" concept question.

As part of this movement, thousands of ConcepTests are available at an online platform known as the Concept Warehouse [5]. The ConcepTests provide multiple choice questions that typically do not require calculations. But they also provide the opportunity for students to enter written explanations for their answers, which is extremely valuable for student metacognition [6], [7], [8], and instructor ability to gain insight into student reasoning [9], [10], [11]. Prior research has reported on results of examining written explanations, with the general finding that even when students answer the basic question "correctly," they often do not provide adequate justification [12], [13], [14].

In this study, the following research questions are posed and investigated:

R1. How do faculty develop, debate, and refine phrasing of concept questions?
R2. How do student responses depend on question phrasing?

To address these questions, a case study is presented of the iterative development of a specific concept question dealing with the topic of rigid body equilibrium. Over a period of approximately three years, the question phrasing was debated and modified amongst a group of interested faculty in a Community of Practice (CoP), with the goal to steer students to address the essential underlying concept, yet in a manner that is not too suggestive of the correct answer. Indeed, prior research indicates that question wording [15] and "surface features" [16] can greatly influence where students place their attention in answering questions. The results of this study suggest that changes of wording can have tangible effects on how students interpret and respond to questions, particularly the degree to which they address the key underlying issue. Therefore, it is important for instructors to use deliberate language in developing, deploying, and discussing such questions.

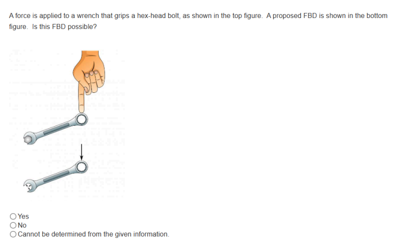
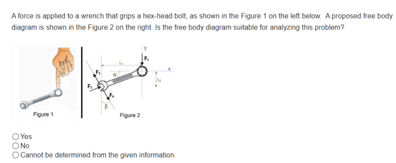
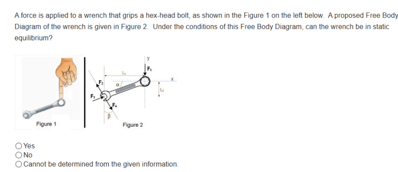
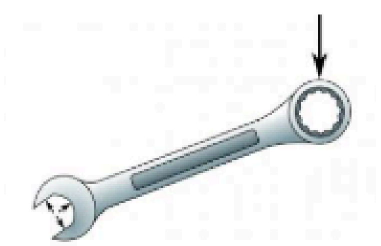
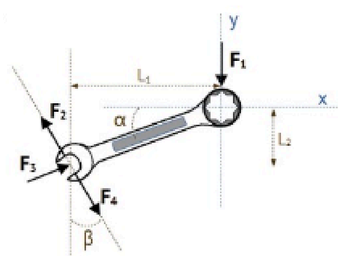
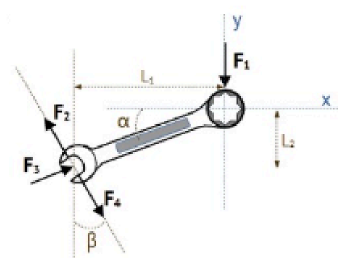
2. Methods

The Concept Warehouse's ConceptTest under examination can is hereafter referred to as the "Wrench Problem" (see Table 1). In this problem, a wrench acts on a hex-bolt, and in turn, the external force applied to the end of the wrench handle is transmitted by a single finger, so as to impart a vertical force with no applied torque (that is, the hand is not 'gripping' the handle). A proposed free body diagram of the wrench is given, showing three concurrent reaction (contact) forces of the bolt on the wrench. The essential question is whether the wrench, as proposed in the FBD, can be in static equilibrium. The underlying intention is that the students will notice the concurrency of the reaction forces, consider/apply the sum of moments about that point, and conclude that the force applied by the finger causes an unbalanced moment, i.e., "No", the wrench cannot be in equilibrium. Table 1 summarizes the three different versions (iterations) of the "Wrench Problem" that were created to elicit this response through the discussions in the Community of Practice (CoP).

To address R1, " How do faculty develop, debate, and refine phrasing of concept questions?", an account of the history of the problem development will be provided. This narrative represents the authors, many of whom were directly involved in the various conversations, and who otherwise represent the active members of the CoP.

To address R2, the complete history of the deployment of all three versions was queried by the platform manager. This data includes the name of each instructor who deployed the problem, an anonymous student ID number of each respondent, the date of the response, and the response details. Response details include the respondent's answer to the closed form question (*Yes*, *No, Cannot be determined from the given information*), a written explanation of their reasoning, and ratings of their confidence in the correctness of their answer, and of their perception of the question's clarity.

Table 1. Versions of the “Wrench Problem”.

Version 1 **(first deployed November 17, 2021)**	**Version 2** **(first deployed November 20, 2023)**	**Version 3** **(first deployed October 15, 2024)**
		
	 Figure 2	 Figure 2
A force is applied to a wrench that grips a hex-head bolt, as shown. A proposed FBD is shown. **Is the FBD possible?**	A force is applied to a wrench that grips a hex-head bolt, as shown in Figure 1 on the left. A proposed free body diagram is shown in Figure 2. **Is the free body diagram suitable for analyzing this problem?**	A force is applied to a wrench as shown in Figure 1. A proposed Free Body Diagram of the wrench is given in Figure 2. Under the conditions of this Free Body Diagram, **can the wrench be in static equilibrium?**
For each version, the allowable responses are **“Yes”**, **“No”**, and **“Cannot be determined from the given information”**		
Notes: Row 1 contains screenshots as seen by students. Row 2 shows enlargement of the FBD portion of the figures, for clarity. Row 3 provides the text of the problem. Note that in all cases, the reaction forces that the bolt applies to the mouth of the wrench are drawn to be concurrent through the bolt center. The essential variations in phrasing are **underlined in boldface** for quick reference.		

The primary method of this article is the analysis of the written explanations provided, across the three versions, to estimate the degree to which the student appeared to address the essence of the problem, which is the lack of moment equilibrium in the proposed FBD. The explanations were searched for keywords of two categories: those that target “moment equilibrium” (ME), and those that target “kinematic parameters” (KP). The ME keywords suggest an attempt to address moment equilibrium such as via the concurrency of the reactions at the mouth of the wrench and the applied moment of the finger. KP keywords suggest a distraction on the parameters at the expense of thinking about moment equilibrium. Several key words and word roots were selected to target the identification of each type of explanation, as shown in Table 2. The intention is that as the number of incidences of keywords increases in a given explanation, the more likely that explanation is to identify the essence of the explanation.

Table 2. Key words and word roots to identify Helpful and Unhelpful Explanations.

Words and roots for ME Explanations	Words and roots for KP Explanations
moment, torq, equilib, balanc, zero, contact, cent (truncation of *center*, *centered* or *centralized*), *concur*	*Dimension, detail, coord, ax* (truncation for *axis* or *axes*), *angl, distance, value*

As a clarifying note, keywords such as 'force' and 'wrench', which appear frequently in the explanations, were omitted. This is based on experience which suggests that such keywords do not necessarily, by themselves, help students approach the essential issue, and so they are not directly helpful to identify either explanations that focus on the essential concept of moment equilibrium. At the same time, they do not directly address the potential distraction of the geometrical parameters. Several exemplar responses – correct and incorrect – that were "tagged" by the keyword search are shown in Table 3. Note that misspellings and other abbreviations (students may use computer, tablet or smartphone) are not caught in this search (e.g., 'cord' may be typed for 'coordinate'; 'equil' for 'equilibrium').

3. Results

The results for R1 are presented as a historical narrative of the development of the Wrench Problem. One author created a similar problem as part of a final exam over a decade ago. When they joined the CoP, they adapted the question for the online platform, as shown in Table1/Version 1. After a year or two of its deployment, other members of the CoP raised questions, such as whether students would interpret the word "possible" as conforming to established standards or rules. In particular, because many instructors consider kinematic dimensions and parameters as a requirement, a concern was raised, that the absence of any dimensions lead some students to answer "No", without considering the "true" essence of the problem. In the words of one colleague, "no axis = wrong" is a mentality that many students bring. Another concern was whether it was clear that the contact reactions were concurrent, particularly without any lines of reference drawn. A third concern was that the word "possible" was ambiguous; i.e., a "possible" FBD need not be a "correct" FBD.

With this in mind, the second version of the problem was created, principally developed by two other authors. In this version two key changes were made: first, dimensions, coordinate axes, and force labels were added to the figure; second, the question was reworded to ask if the FBD was 'suitable'; i.e., a correct FBD that is able to be used to used to analyze the system.

After a year of deployment of this version, another discussion followed, returning to some of the same issues. The 'flip' of the prior concern was now advanced, that some students now focus on the presence of dimensions, again bypassing the (intended) essence of the problem. Also, the word “suitable” was questioned for possible ambiguity.

With this in mind, a third version of the question was advanced, principally by the author who developed Version 1. This new Version 3 maintained the diagram from Version 2, due to the group consensus that incorporating dimensions in the FBD is a best practice. However, instead of asking if the FBD is ‘possible’ or ‘suitable for analysis’, the question now asks, somewhat more directly, if the wrench can be in equilibrium.

To address R2, the complete deployment history for each version of the Wrench Problem was conducted, including the closed form responses, written explanations, and student ratings for clarity. Then, each explanation was evaluated for the number of incidences of the ME and KP keywords. For insight into the process, Table 3 contains some examples (from all versions of the question) of how the keyword incidences were counted, and how well they foretold the anticipated result.

Table 3. Examples of tabulations of keywords from student explanations.

Ref ID	Answer Submitted	Explanation (identified keywords are underlined)	Number of Incidences and type	Commentary
159422	*No* (correct)	Because the sum of the **moment**s around the **cent**er of the hex-head bolt does not equal zero, this FBD is not possible	2 ME	This is an excellent response, which was accurately detected by the 2 ME keywords
158542	*No* (correct)	Assuming the wrench is in **equilib**rium, the force cannot exist as it will cause a **moment** at the bolt.	2 ME	This is a very good response, which was accurately detected by the 2 ME keywords (it is slightly ambiguous as to what the student meant by “the force”).
159434	*Yes* (incorrect)	This FBD is possible as the correct forces are represented and these forces could be **balanc**ed to **zero** (the system could be at **equilib**rium).	3 ME	This is an incorrect response, despite the positive indicator of 3 ME keyword instances.
157771	*No* (correct)	There are four unknown forces but only three possible **equilib**rium equations (net forces in the x-direction, net forces in the y-direction, and net **moment** about a point), therefor [sic] this FBD cannot be solved.	2 ME	The submitted answer is correct, but the explanation is not. Nevertheless, the explanation has procedural merit, and so the 2 ME keyword instances are useful.
155555	*No* (correct)	No I do not think that the given FBD is possible because there are 4 unknowns shown in the diagram, and we can only use 3 equations, which will not allow us to solve for all of the unknowns.	0	This answer and response are essentially similar to the previous response from ID157771. However, the explanation was written without the presence of any of the keywords.
153510	*Yes* (incorrect)	this fbd works because it displays all of the necessary forces	0	This response is typical of many that does not contain any keywords, as it

				neither addresses the moment equilibrium, nor the surface features.
163368	*Yes* (incorrect)	This free body is affective [sic] because it has all of the proper forces and applied forces being acted upon this wrench. The free body did not include a hex-head bolt, and only the wrench that is applying a force to this bolt. There are the correct number of lengths and distances that are required to take sum of **moment**s, also including the required **angl**es. The only inconvenient part of this free body is the **ax**is being at one loop of the wrench, which would be more convenient including the whole free-body.	2 KP 1 ME	The presence of 2 KP comments is useful to detect responses such as this which emphasize the need for specific parameters in order to 'solve' the problem; this illustrates a confusion between 'solving' and 'being able to determine'.
158801	*No* (correct)	The proposed free body diagram does contain appropriate **angl**es, and **dimension**s so the sticking point if there is a sticking point is in the forces involved. F1 is clearly a given as is f2, and f4; the justification for F3 is a little bit more suspect given the forces involved push the wrench directly downwards and around the wrench rather than into it.	2 KP	The presence of 2 KP comments is again useful to detect responses that miss the essential idea in the problem.

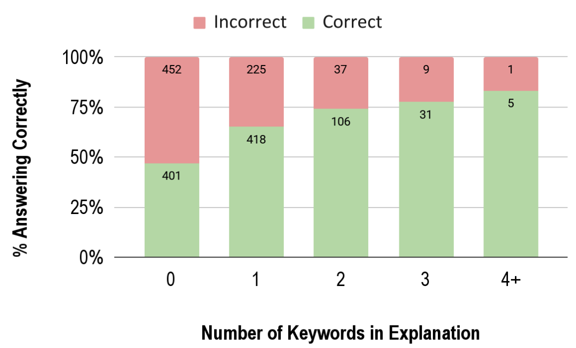
To date, the wrench problem has a total of 2,079 known deployments, 1,386 of which were conducted in the Concept Warehouse platform, with the remaining 693 being borrowed and conducted through another platform. For the purposes of this study, only the 1,685 responses (from both platforms) which have written explanations are evaluated (557, 881, and 247 for Versions 1, 2, and 3, respectively). Analysis of this data suggests that the presence of multiple ME keywords are useful to identify explanations that address moment equilibrium, and that the presence of multiple KP keywords tend to identify explanations that avoid addressing moment equilibrium.

As shown in Table 4, as the number of ME keywords increases, the more likely it is that a student will select the correct answer "No". Although previous research has demonstrated that the selection of a correct response does not guarantee that a sufficiently correct or complete explanation is provided [12], [13], [14], the results here show something similar to the converse: as the presence of ME keywords increases, so does focus on the correct reasoning, which often leads to the correct answer. Also as shown in Table 4, as the number of KP keywords increases, the less likely it is that students select the correct answer.

Table 4. Tabulation of Keywords, Demonstrating Correspondence with Identification of Intended Content.

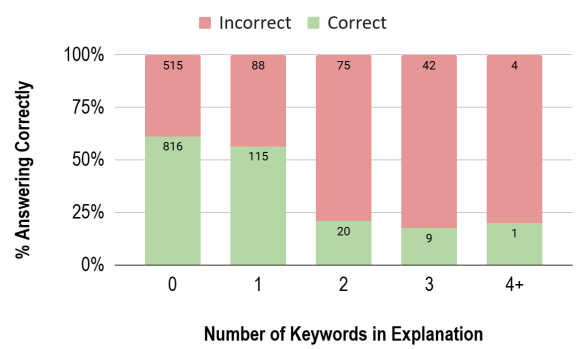
Tabulation of ME Keywords, all versions

	Correct	Incorrect	Total	%Correct
0 keywords	401	452	853	47.0%
1 keyword	418	225	643	65.0%
2 keywords	106	37	143	74.1%
3 keywords	31	9	40	77.5%
4+ keywords	5	1	6	83.3%
Total	961	724	1685	57.0%



Tabulation of KP Keywords, all versions

	Correct	Incorrect	Total	%Correct
0 keywords	816	515	1331	61.3%
1 keyword	115	88	203	56.7%
2 keywords	20	75	95	21.1%
3 keywords	9	42	51	17.6%
4+ keywords	1	4	5	20.0%
Total	961	724	1685	57.0%



To compare the effect of question phrasing, the subset of the data (N = 544) is examined that corresponds to four instructors (denoted "Group A") who deployed each of the three versions of the problem. This is useful because this controls for possible effects of teaching style (that is, each instructor uses a similar method of teaching and deployment of the problem across all different versions). The results, which are summarized in Table 5, appear to indicate that data for Versions 1 and 2 have similar profiles with respect to the ME keywords, and Versions 1 and 3 have similar profiles for the KP keywords.

As seen in Table 6, the data suggest that the effect of using the phrase "can the wrench be in static equilibrium?" in Version 3 was very effective to focus students' attention on the underlying issue of moment equilibrium. Comparison of the two profiles (Versions 1+2 vs Version 3) using the Mann Whitney U Test yields $p < 0.0001$. This suggests that the presence of the word "equilibrium" in the question phrasing triggers focus on that word. The number of correct responses for Version 3 is also higher (51.7%) than for the aggregated Version 1+2 (43.0%).

The situation for the KP words is not quite as intuitive. The data in Table 7 indicate that the aggregated profile from Versions 1+3 is significantly different from the profile of Version 2, again due to the Mann Whitney U Test yielding $p < 0.0001$. This suggests that the effect of

Version 2, where dimensions were first added to the figure, was to draw attention to the details of the dimensions, and possibly distract from the essential issue of moment equilibrium. Yet because Version 3 has the identical figure as Version 2, the kinematic details do not necessarily dominate the thought process when other triggers, such as the word "equilibrium", are present.

Table 5. Summary of responses for Group A.

	Correct and Total Responses (% correct)	Number of Keyword Incidences	Number of ME Keywords in Explanation	Number of KP Keywords in Explanation
Version 1	80/203 (39.4%)	0	132 (65.0%)	182 (89.7%)
		1	60 (29.6%)	14 (6.9%)
		2	10 (4.9.%)	6 (3.0%)
		3	0 (0.0%)	1 (0.5%)
		4+	1 (0.5%)	0 (0.0%)
Version 2	65/134 (48.5%)	0	93 (69.4%)	100 (74.6%)
		1	35 (26.1%)	26 (19.4%)
		2	6 (4.5%)	7 (5.2%)
		3	0 (0.0%)	1 (0.7%)
		4+	0 (0.0%)	0 (0.0%)
Version 3	107/207 (51.7%)	0	51 (24.6%)	181 (87.4%)
		1	94 (45.4%)	22 (10.6%)
		2	39 (18.8%)	3 (1.4%)
		3	19 (9.2%)	0 (0.0%)
		4+	4 (1.9%)	1 (0.5%)

Table 6. Distributions of ME Keywords in Explanations from Group A.

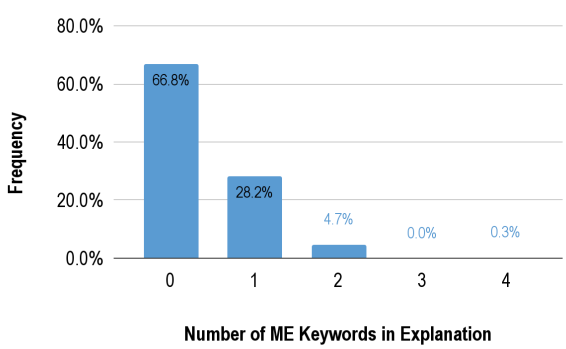
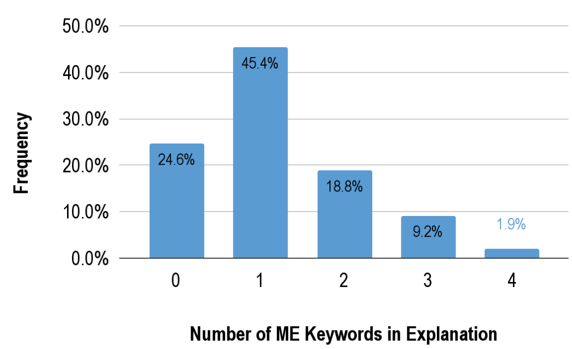
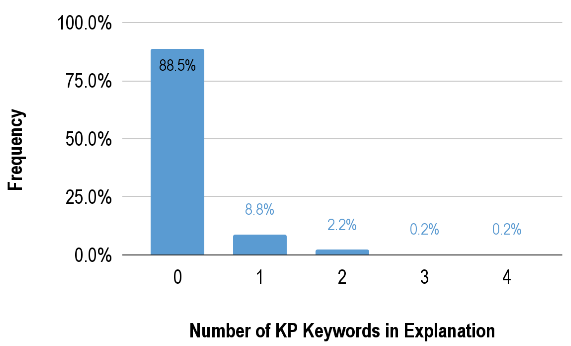
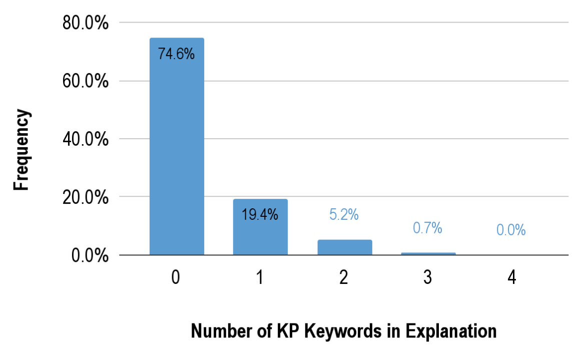
Versions 1 and 2 Aggregated Correct/Total = 145/337 (43.0%)	Version 3 Correct/Total = 107/207 (51.7%)
	
Mann Whitney U Test: $p < 0.0001$.	

Table 7. Distributions of KP Keywords in Explanations from Group A.

Versions 1 and 3 aggregated **Correct/Total = 187/410 (45.6%)**	**Version 2** **Correct/Total = 65/134 (48.7%)**
	
Mann Whitney U Test: $p < 0.0001$.	

Having said this, in the development of this article, the authors realized that teaching style might have as significant an influence on the outcome of the question as the phrasing itself. Therefore, another grouping emerged of instructors who “do” (Group B) and those who “do not” (Group C) emphasize or require kinematic markers (axes, dimensions, angles, etc.) to be drawn on FBD’s. Group B is further subdivided into Group B1, of instructors deploying the problem on the Concept Warehouse platform, and Group B2, of an instructor using an alternative platform. The results of this analysis are provided in Tables 8 and 9. A visual inspection of the graphs in these tables shows two principal trends:

- For a given instructional approach, Version 3 elicited the highest incidence of ME keywords, as well as the highest correct response rate for “No”.
- For a given instructional approach, Version 2 elicited the highest incidence of KP keywords, as well as higher correct response rates compared to Version 1.

These results seem to corroborate the results from Group A, indicating that the direct inclusion of the word “equilibrium” is likely to focus students’ attention on that issue, making it more likely that they will address the underlying issue of moment equilibrium. The results also suggest that the presence of geometrical parameters in the diagram can possibly distract students attention away from the underlying issue, especially in the absence of other “helping” keywords that bring attention back to the main issue. On the other hand, while commenting on dimensions, axes, etc., may be part of the students’ attempts to articulate their reasoning, the change in figure (and question) may have provided more clarity to the *Yes/No* question.

Table 8. Distributions of ME Keywords for Groups B1, B2, and C.

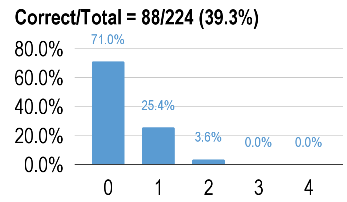
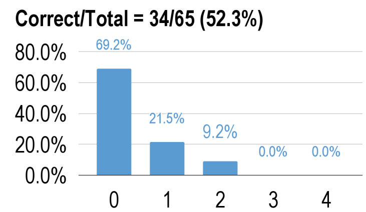
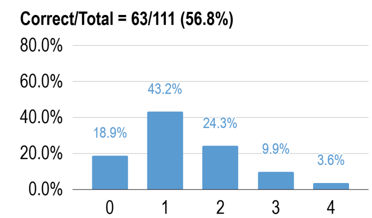
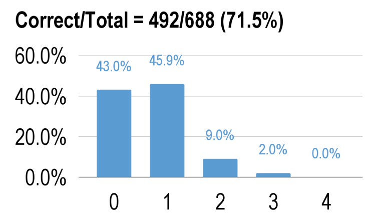
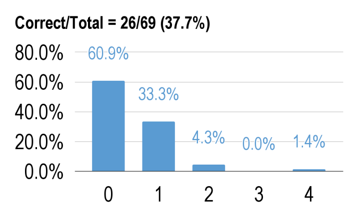
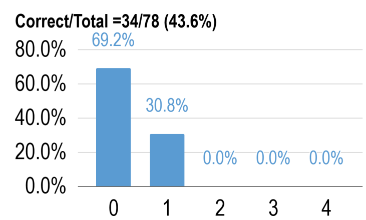
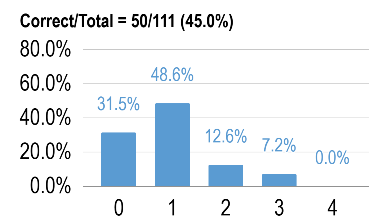
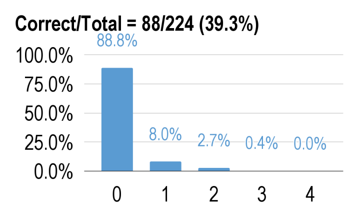
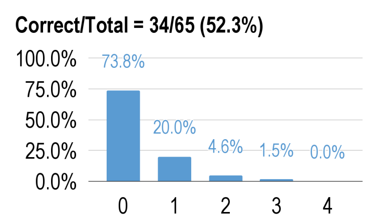
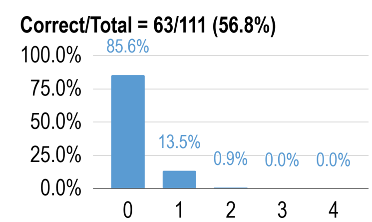
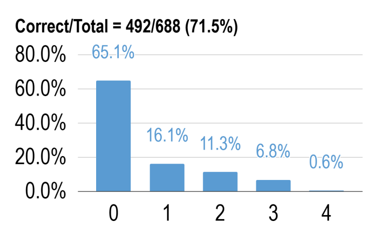
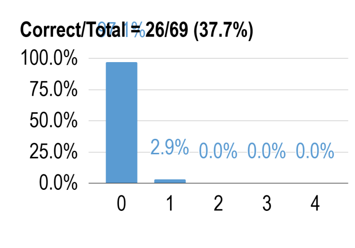
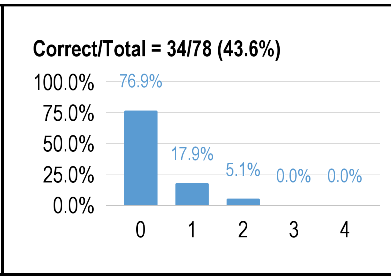
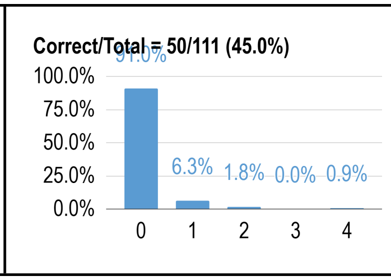
Gr	Version 1	Version 2	Version 3
B1			
B2	n/a		n/a
C			

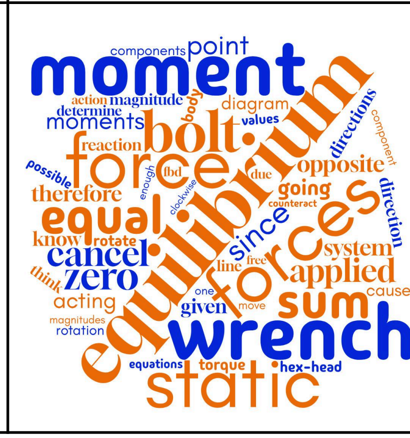
Table 9. Distribution of KP Keywords for Groups B1, B2, and C.

	Version 1	Version 2	Version 3
B1			
B2	n/a		n/a

C			

As a final comment, the concept warehouse platform generates word clouds to visualize student explanations [17], and presented in Table 10. Notably, the word cloud for Version 3 prominently displays "moment" and "equilibrium" with high frequency, in correspondence with the tabulated results. The word "angles" also figures prominently for Version 2, also in correspondence to the tabulated results. The word "force" appears frequently in all three word clouds, but this word was deliberately not tracked in the analysis, due to the sense that force by itself is neither directly helpful to understanding the underlying issue of moment equilibrium (force equilibrium is not the issue), nor is it unhelpful in the sense of looking for surface features.

Table 10. Word Clouds generated from Student Explanations.

Version 1	Version 2	Version 3
		

4. Discussion

Making an effective concept question is an art. When rephrasing a question to address one ambiguity, it is easy to create another ambiguity which needs to be addressed. The writer can quickly turn into the proverbial dog chasing its tail–never catching the perfect question phrasing.

Question phrasing depends on the specific context in which each instructor uses the question and on instructor preferences. Writing a problem within a CoP has advantages and disadvantages. On the positive side, the different viewpoints of each member and different contexts within which they use questions provides a broad understanding of how a question can be interpreted. This leads, in general, to more robust questions. On the downside, the process is time consuming and individual instructor preferences cannot always be resolved. For example, this CoP could not agree on what dimensions, measurements, and angles were needed in a "proper" free body diagram. Writing questions for use by a larger community of users outside the CoP is an even greater challenge. It might not be possible to write a single 'neutral' version of a problem that will be interpreted 'equitably' among all students across a broad range of instructor preferences and styles.

Stepping back, the three versions may be considered snapshots along the path of creating and using a FBD. While instructors (experts) may see this as one continuous process, students may not see the interplay between picture and application, and may get bogged down in details. The versions incrementally step from the FBD with only unlabeled forces, to detailing the FBD with dimensions, to the verge of applying equations by explicitly asking about "equilibrium". In Version 1, the FBD includes only the potential forces. Do some students just think about the picture itself (right or wrong) without worrying about how the forces are generated and/or how they affect equilibrium? In Version 2, dimensioning the FBD implies moment equilibrium, but while this is a statics class, is not the point of a wrench to turn a bolt? In the prerequisite physics course, most problems with FBDs had something to do with acceleration, not equilibrium. Do students understand "suitable" to mean "correct to solve the equilibrium problem"? While forces are now labeled with variables, do students take all four as "unknowns" and thus conclude the system is statically indeterminate, as in two example responses in Table 3? In Version 3, the drawing part of the FBD is "done" and the question focuses on the next step – applying equilibrium. Is this last version sufficient to ensure students are responding to the underlying concepts that the authors believe they are asking about?

Leading students to focus on key details, without over-describing the problem, so that the detail is 'given away' is difficult. For example, the inclusion of "equilibrium" in the problem phrasing of Version 3 seemed to elicit student attention to that issue. However, their attention did not necessarily extend to considering moment equilibrium and the concentric/concurrent nature of the reactions. Including the phrase "moment equilibrium" may have improved student responses to the question, but at the cost of reinforcing the need for students to identify critical details of a problem.

5. Conclusions

There are at least two important conclusions of this work. First, it is clear the problem phrasing does have an impact on how students respond, and moreover, well-crafted phrasing can elicit responses to essential issues. In this case study, the explicit focus of the word "equilibrium" in the problem statement led students to shift their attention toward the issue of moment equilibrium. Second, it is clear that faculty collaboration, such as through a community of practice, can lead to such well-crafted problems. Future extensions of this work can inquire about differences among gender, and possibly further experiments with other variations of the problem phrasing and image presentation.

Acknowledgments

This material is based upon work supported by the National Science Foundation under a collaborative grant with grant numbers DUE 1821638, 1821439, 1821445, 1820888, 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.

Bibliography

[1] I. D. Beatty, W. J. Gerace, W. J. Leonard, and R. J. Dufresne, "Designing effective questions for classroom response system teaching," *Am. J. Phys.*, vol. 74, no. 1, pp. 31–39, 2006, doi: 10.1119/1.2121753.

[2] C. Papadopoulos, A. I. S. Roman, M. J. Perez-Vargas, G. Portela-Gauthier, and W. C. Phanord, "Development of an alternative statics concept inventory usable as a pretest," in *ASEE Annual Conference and Exposition*, 2016.

[3] P. S. Steif and M. A. Hansen, "New Practices for Administering and Analyzing the Results of Concept Inventories," *J. Eng. Educ.*, no. July, pp. 205–212, 2007, doi: 10.1002/j.2168-9830.2007.tb00930.x.

[4] G. L. Gray, F. Costanzo, D. Evans, P. Cornwell, B. Self, and J. L. Lane, "The Dynamics Concept Inventory Assessment Test : A Progress Report and Some Results Introduction," in *Proceedings of the ASEE Annual Conference and Exposition*, 2005, p. 10.1278.1-10.1278.16. [Online]. Available: https://peer.asee.org/15378

[5] M. D. Koretsky *et al.*, "The AiChE Concept warehouse: A web-Based tool to Promote Concept-Based instruction," *Adv. Eng. Educ.*, vol. 4, no. 1, 2014, Accessed: May 31, 2021. [Online]. Available: https://eric.ed.gov/?id=EJ1076138

[6] P. M. Papadopoulos, N. Obwegeser, and A. Weinberger, "Concurrent and retrospective metacognitive judgements as feedback in audience response systems: Impact on performance and self-assessment accuracy," *Comput. Educ. Open*, vol. 2, p. 100046, Dec. 2021, doi: 10.1016/j.caeo.2021.100046.

[7] "Why does peer instruction benefit student learning? | Cognitive Research: Principles and Implications | Full Text." Accessed: Jan. 15, 2025. [Online]. Available: https://cognitiveresearchjournal.springeropen.com/articles/10.1186/s41235-020-00218-5

[8] L. S. Grundy and M. Koretsky, "Student Metacognitive Reflection on a Conceptual Statics Question," presented at the 2023 ASEE Annual Conference & Exposition, Jun. 2023. Accessed: Jan. 15, 2025. [Online]. Available:

https://peer.asee.org/student-metacognitive-reflection-on-a-conceptual-statics-question
[9] M. Carlson, M. Oehrtman, and N. Engelke, “The Precalculus Concept Assessment: A Tool for Assessing Students’ Reasoning Abilities and Understandings,” *Cogn. Instr.*, vol. 28, no. 2, pp. 113–145, Apr. 2010, doi: 10.1080/07370001003676587.
[10] W. Hung and D. Jonassen, “Conceptual understanding of causal reasoning in physics,” *Int. J. Sci. Educ.*, vol. 28, no. 13, pp. 1601–1621, 2006.
[11] C. Igaz and M. Prokša, “Conceptual Questions and Lack of Formal Reasoning: Are They Mutually Exclusive?,” *J. Chem. Educ.*, vol. 89, no. 10, pp. 1243–1248, Sep. 2012, doi: 10.1021/ed100895c.
[12] M. D. Koretsky, B. J. Brooks, and A. Z. Higgins, “Written justifications to multiple-choice concept questions during active learning in class,” *Int. J. Sci. Educ.*, vol. 38, no. 11, pp. 1747–1765, 2016, doi: 10.1080/09500693.2016.1214303.
[13] C. Papadopoulos, E. Davishahl, C. H. Ramming, J. C. B. Abreu, and W. A. Kitch, “Work in Progress: Context Matters: A Comparative Study of Results of Common Concept Questions in Statics at Several Diverse Institutions,” *ASEE Annu. Conf. Expo. Conf. Proc.*, 2022.
[14] C. Papadopoulos *et al.*, “Context Matters: Continued Study of Results of Common Concept Questions at Several Diverse Institutions,” presented at the 2023 ASEE Annual Conference & Exposition, Jun. 2023. Accessed: Jan. 15, 2025. [Online]. Available: https://peer.asee.org/context-matters-continued-study-of-results-of-common-concept-questions-at-several-diverse-institutions
[15] M. A. McDaniel, J. L. Anderson, M. H. Derbish, and N. Morrisette, “Testing the testing effect in the classroom,” *Eur. J. Cogn. Psychol.*, vol. 19, no. 4–5, pp. 494–513, Jul. 2007, doi: 10.1080/09541440701326154.
[16] M. Weston, K. C. Haudek, L. Prevost, M. Urban-Lurain, and J. Merrill, “Examining the Impact of Question Surface Features on Students’ Answers to Constructed-Response Questions on Photosynthesis,” *CBE—Life Sci. Educ.*, vol. 14, no. 2, p. ar19, Jun. 2015, doi: 10.1187/cbe.14-07-0110.
[17] B. Brooks, D. M. Gilbuena, S. Krause, and M. Koretsky, “Using Word Clouds for Fast, Formative Assessment of Students’ Short Written Responses,” *Chem. Eng. Educ.*, Sep. 2014, Accessed: Jan. 15, 2025. [Online]. Available: https://www.semanticscholar.org/paper/Using-Word-Clouds-for-Fast%2C-Formative-Assessment-of-Brooks-Gilbuena/d41ae27d49e10aa7ea2a8496cc66e506c2ff0ccb