Work in Progress for Two Questions: Confusion Matrix Analysis of Student Think-Alouds during a Dynamics Concept Inventory Exam

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 (USI). He loves teaching all of the basic mechanics courses, and of course his Vibrations and Finite Element Analysis courses.

Dr. Andrew Jason Hill, University of Southern Indiana

Jason Hill is an associate professor of engineering and director of civil engineering at the University of Southern Indiana. He holds B.S., M.S., and Ph.D. degrees in civil engineering from Tennessee Technological University. His research interests include rainfall-runoff modeling, stream restoration, and wetland hydrology.

Confusion Matrix Analysis of Student Think-Alouds during a Dynamics Concept Inventory Exam: Work in Progress for Two Questions

Julian L. Davis & A. Jason Hill

Concept inventory question results are frequently used to evaluate students' overall understanding of course material. These questions focus only on foundational concepts rather than detailed problem solving. Dynamics Concept Inventory (DCI) exam results were analyzed for two semesters. Students were asked to audibilize their thoughts while taking the exam and being recorded. The video recordings of students taking the exam were transcribed for two questions. Results of two independent analyses of the transcriptions were summarized statistically in a confusion matrix. The confusion matrix is a 2x2 matrix with column headings scoring student oral reasoning and row headings scoring question outcome. These results are expected to illuminate the influence of luck versus true understanding during the exam.

Introduction

Concept inventories have been around for many years now. As their title suggests, they test one's understanding of concepts, and require no calculations. They can be considered as a way to understand a person's thinking process. Concept inventories have been developed for Physics, [1], Statics [2], [3], Dynamics [4] , Strength of Materials [5], Heat Transfer [6] and many other statistics, engineering and/or physics topics [7].

Faculty often say that we want students to understand the concepts behind what we teach – to see a bigger picture. But many also feel students need to understand how to problem solve as well. And how many times have we heard from students that they just want to know how to solve "THIS TYPE OF PROBLEM." So, there is often a tug of war between these two ideas in the classroom. In some cases, there can be a mismatch between what students want and what faculty want their students to be able to do.

Concept inventories are constructed very carefully with answers to help infer students' thoughts based on the answers that they give. There are many different assessment and development methods [8]. But wouldn't it be nice just to sit inside a person's head while they are taking the concept exam and listen to what they are thinking?

Well, we tried just that, although we were not ACTUALLY INSIDE their head. We asked students to audibilize their thoughts in a non-threatening environment. However, while taking the exam, the students audio and video were recorded for future transcription.

Methods

With the permission of the authors of the original Dynamics Concept inventory (DCI), we administered the DCI online (as opposed to using a printed paper copy). A proctoring software was used to record audio and video of students taking the exam. Students were asked to audiblize their thoughts while taking the exam. Because the videos were not easily downloadable, they were re-recorded using Zoom. Audio during the concept inventory exam were transcribed by the lead author. The results were evaluated independently by both authors for correctness and oral reasoning.

Only two questions from the DCI were selected for this initial analysis. The question regarding Coriolis acceleration (Question 5) was selected because it was the problem that first spawned the idea for this paper. Question 22, on No-Slip velocity was selected due to its variety in selected answers in the exam. The question statements as given to students are shown in Figure 1.

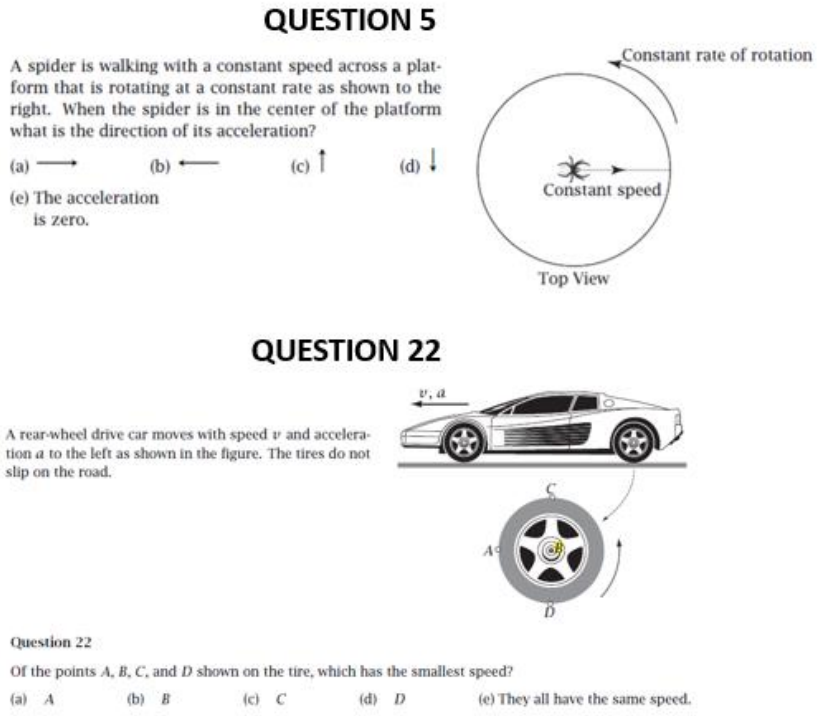

Figure 1. Two questions (5 and 22) from a Dynamics Concept Inventory Exam

Scores were classified into one of four categories using a confusion matrix (Table 1). A confusion matrix is a commonly used statistical classification tool that is a 2x2 contingency table [9]. The matrix can provide insight to the types of errors students are making. In this study, the confusion matrix provides a convenient way to score question outcome (correct or incorrect) on the concept inventory exam regarding student oral reasoning, as reflected in the transcribed student "think-alouds". Of course, the ideal outcome is a correct response with the correct oral reasoning (i.e., correct for the right reasons). A correct answer can also occur with flawed oral reasoning. These two outcomes are the top row of the confusion matrix in Table 1. If a student answers incorrectly, the student may have correct or incorrect reasoning. These two outcomes are the second row of the confusion matrix. The first column total reflects correct student oral reasoning regardless of student outcome. Finally, if a student provided no reasoning at all (they were silent during the question), they were provided a score of incorrect reasoning. Statistics from confusion matrix results are expected to be insightful.

Table 1. Confusion matrix used to score concept inventory student transcriptions regarding oral reasoning.

		Student Oral Reasoning	
		Correct	Incorrect
Question Outcome	Correct	Correct Answer and Correct Reasoning	Correct Answer but Incorrect reasoning
	Incorrect	Incorrect Answer but Correct Reasoning	Incorrect Answer and Incorrect reasoning

Results

The confusion matrix results are shown in Figure 2 for both concept inventory questions and independent evaluations. Each question has 4 matrices, pre- and post-course for each evaluator. Overall performance on both questions 5 and 22 was poor as evidenced by the large number of incorrect responses who also had incorrect reasoning (row 2 and column 2 of each matrix). Student performance did not improve post-course for Question 5 and only marginally improved for Question 22. Independent evaluations by the authors yielded an identical classification for Question 22 and a minor difference for Question 5.

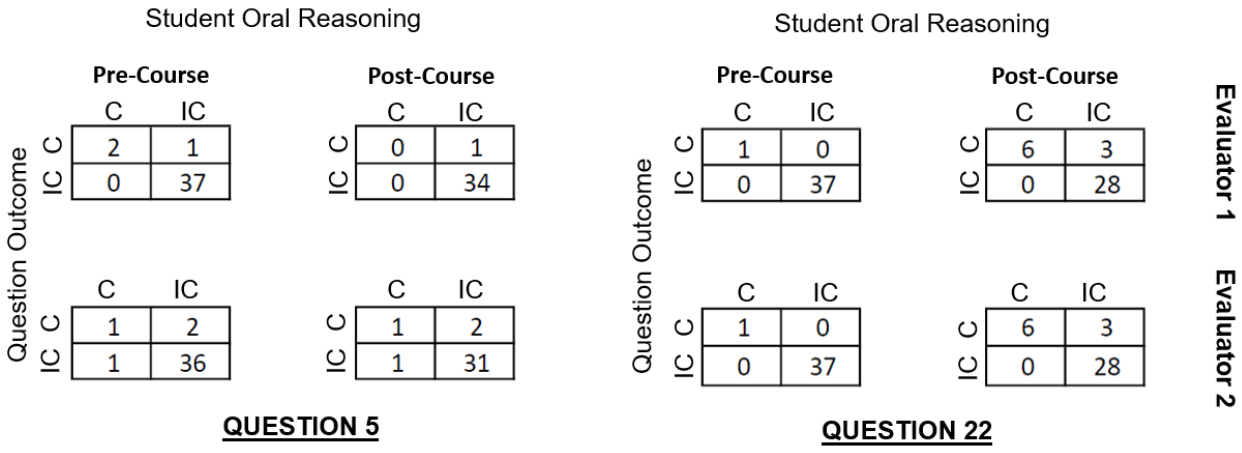

Figure 2. Confusion matrix results summarized for Question 5 and 22 as scored by independently by the authors (Evaluator 1 and 2).

Additional insight into the student thought process comes from the transcriptions. A few examples are shown below:

Question 22 Transcription Examples (Pre-Course):

Incorrect Responses

- The smallest V. They're all the same … in which direction. I'm going to say they all have the same speed. Unless it's D. But I think its … just like, they're all moving … the whole car … you can't have different velocities on one part of the car.
- The smallest speed would be B in this one because B is toward the center and A C and D will have the same.
- Which has the smallest speed. A B C or D … wow this is what got me in Physics, I really need to review this. Um I think A C and D speed … but B … No they might just all have the same speed if we're looking at … yeah … I'm going to say they have … I'm going to say E
- The "S" and "L" component of velocity based on the tire's movement. Since it's in the very center basically everyone one of them is moving at B, the velocity of the car, but each has their own velocity of the tire s well, except for B.

Correct Responses

- Nothing was said aloud – scored as incorrect reasoning

Question 22 Transcription Examples (Post -Course):

Incorrect Responses

- Ehh … of the points A … shown … which one has the smallest speed. Well … don't they all have the same speed. Wait … I would like say that they all have the same speed … which … I mean I know that there's not acceleration at D, but … that doesn't necessarily mean that they all don't have the same speed. So, I like to say they all have the same speed.
- Because we're being asked about the velocity not the angular velocity. The velocity is going to be the same for all of the points.

Correct Responses

- So this one since D is touching the ground it has zero velocity.
- Which has the small speed. So B, because it only has the V, No it also has the velocity plus B. If it was the velocity at this point, then it would be D, because It's going opposite. it's counter acting V. We know D wouldn't be moving. No … it wouldn't.
- Of the points ABC and D, shown on the tire, which has the smallest Speed. Um smallest speed will be D, because it's just touching the ground, there's no slip. Um D will be Zero, zero speed. So it will be D.

Question 5 Transcription Examples (Pre-Course):

Incorrect Responses

- … So the acceleration will be zero because she's not acceleration because she's in the middle because she's only rotating, but she's not accelerating because she doesn't have a radius around, around, but from the middle there is where the acceleration is coming That's what I guess

- You spider … *reading* … so its acceleration … *reading* … it's at the center … at that instant it would be zero, right? … acceleration when the spider is in the center of the platform what is the direction of acceleration … I want to say zero.

Correct Responses

- … well if it's at the center there's … there's no … no acceleration towards the middle … it's at a constant speed, but it's still have it's still have tangents Tangent, tangential acceleration …Thinking …It'd be A or B … it's C or E … it's in the Center, It still rotating .. so it will still have an acceleration. So it's B. (Note: Student must have said "C".)

Question 5 Transcription Examples (Post-Course):

Incorrect Responses

- Ok this is gonna be the relative thing … *reading* … when it's at the center the direction of the acceleration is zero because it's not moving. As long as it's not backwards cause like that would not make sense. Could be forward cause like that's … but your rotating … but if you're at the center, it's zero. I hope that's right.
- When the spider is in the center of the platform, what is the direction of it's acceleration? So, a spider … So … it starts on the left side. It's … Well, if I start here … it's moving at a constant rate. I mean I need to know … wouldn't I have to know the … Well, I think … when the spider is in the center … Oh acceleration … oh acceleration … oh … we're talking about constant everything … it should be zero … I spent way too much time that I should have.

Correct Responses

- So … it's at the center … so there's no normal acceleration … but there's still tangential acceleration that's spinning it. So that would be, that would be C if it's spinning counterclockwise. But it's a constant speed. Constant …

Discussion

The transcription results indicated multiple students had a misunderstanding of what was happening in both problems. Based on this feedback, the questions were modified as shown in Figure 3 to help clarify some of the issues that students may have run into while reading/answering these questions.

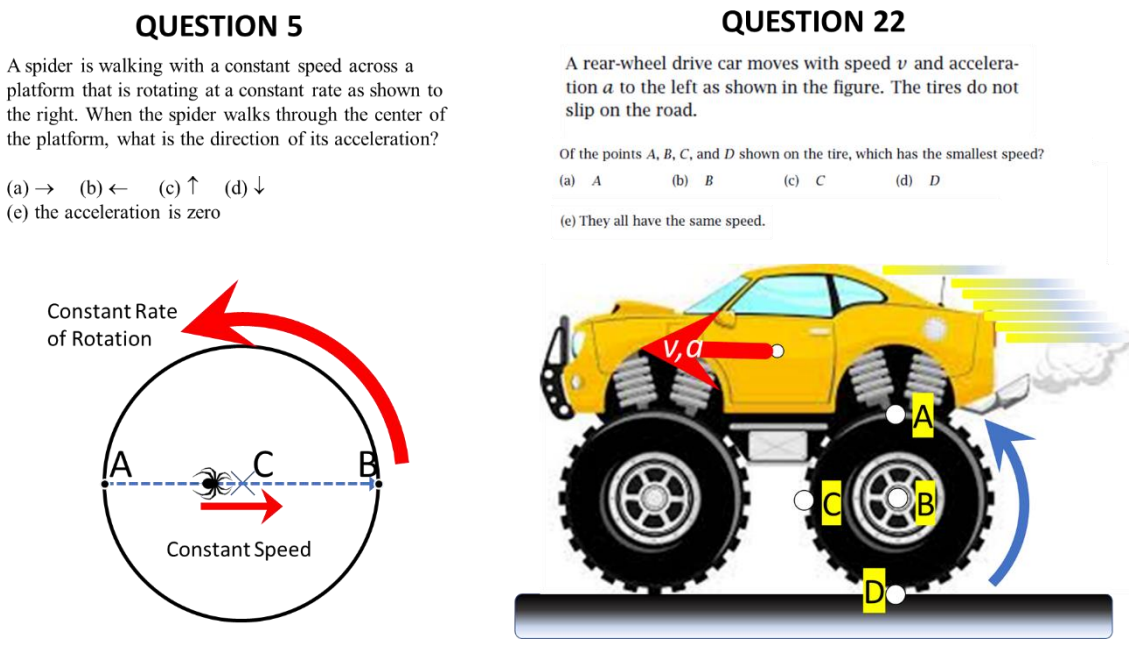

Figure 3. Modified Question 5 and Question 22 problem statements, designed to alleviate student confusion

For Question 5, the key difference compared to the problem in Figure 1 is (1) The spider is no longer positioned at the center to indicate it is not YET at the center of the disk, (2) a dashed line and red arrow is added to illustrate the spiders path started at A and to further emphasize that the spider is moving at a constant speed to the right and (3) the statement says that the spider walks "through the center," hopefully indicating that the spider is not stationary. Reasons for the adjustment are that students seem to think that the spider is already at the middle of the disk and don't understand that it is walking through the center of the disk. They interpret that the spider starts at the middle of the disk.

For Question 22, notice how the wheel is removed from the car in the original question (Figure 1). The image in Figure 3 is similar to the original, but we suggest keeping the wheel attached to the car could help students de-couple this question from a previous one in the DCI. However, we understand that this may not leave enough room for the labeling – so maybe a different vehicle (Monster truck) would provide more room for the labeling. Also, notice that the car is (attempted) to be *shown in motion* by the "speeding" yellow to gray lines.

Though we were not explicitly able to distinguish luck from understanding in our analysis, we think this could be addressed in two ways. First, encouraging more vocalizations would help address this issue. Even though the instructions explicitly asked students to vocalize their thoughts, not all of them did so. Interviews with students could be one effective way to address students who are not vocalizing their thought process during a test and may be a better way of extracting understanding. However, this defeats the purpose of letting the students take the test in a relaxed and comfortable location in which they do not fear vocalizing their thoughts. Secondly, adding a confidence level question to each individual DCI question could allow for a better idea of when students are using luck versus correct understanding to select their answer. The next implementation of the DCI will include a confidence level response.

Finally, students seemed more comfortable taking the exam in silence – which is completely understandable based on how exams are typically administered; but of course, this defeats our purpose of asking students to vocalize their thoughts during the exam. This would be an interesting study in-and-of-itself; giving students an opportunity to talk through a written exam/test may be helpful to identifying solutions for themselves.

Conclusions

Here we analyzed transcriptions of audio recordings of student's vocalizations to two questions while taking the DCI exam. Our intention was to take advantage of the technology where students could 1) manage their own video and audio recording, 2) take the exam in a location comfortable to themselves, and 3) not have outside stressors influence their answers and comfort level while taking the exam. We used a confusion matrix to rate if a student was using correct/incorrect oral reasoning and if it led to the correct/incorrect answer. The results indicate that there is some confusion in students' interpretation of these questions in the DCI. We also have suggestions to adjust images to two DCI images that may help students better answer the concepts that are being tested. This by no means is a suggestion that these changes are the only things that will adjust students' understanding of concepts, but it may help one or two think a little harder about what is happening in the problem. These images have not yet been tested in the DCI.

At this point, we were not able to successfully tease out students selecting the correct answer based on luck vs. clear understanding of the concept tested. In the future, we will include a confidence level rating question for each DCI question. This will allow students to rate their confidence in the answer provided. The combination of 1) confidence level rating along with 2) video/audio recording of the students audibilizing their thought process should be an excellent way to gauge understanding and luck in selecting answers during the DCI exam.

References

[1] D. Hestenes and I. Halloun, “Interpreting the FCI:A Response,” *The Physics Teacher*, vol. 33. pp. 502–506, 1995.

[2] P. S. Steif, “Initial data from a statics concept inventory,” *Proceedings of the 2004 American Society of Engineering Education Conference & Exposition*, 2004, doi: 10.1111/j.1365-2427.2011.02613.x.

[3] P. S. Steif and J. A. Dantzler, “A Static Concept Inventory: Development and Psychometric Analysis,” *Journal of Engineering Education*, vol. 94, no. 4, pp. 1–9, 2013, [Online]. Available: papers3://publication/uuid/BA46CB35-A6E8-41DD-B4AD-49F04BFAE884

[4] G. Gray, F. Costanzo, and D. Evans, “The dynamics concept inventory assessment test: A progress report and some results,” *American Society for Engineering Education*, 2005, [Online]. Available: http://www.esm.psu.edu/dci/papers/ASEE-DCI-Portland.pdf

[5] J. Richardson, P. Steif, J. Morgan, and J. Dantzler, “Development of a concept inventory for strength of materials,” in *33rd Annual Frontiers in Education, 2003. FIE 2003.*, 2003, vol. 1, pp. T3D–T3D.

[6] M. Prince, M. Vigeant, and K. Nottis, “Development of the Heat and Energy Concept Inventory: Preliminary Results on the Prevalence and Persistence of Engineering Students’ Misconceptions,” *Journal of Engineering Education*, vol. 101, no. 3, pp. 412–438, 2012, doi: https://doi.org/10.1002/j.2168-9830.2012.tb00056.x.

[7] D. L. Evans *et al.*, “Progress on concept inventory assessment tools,” in *Proceedings - Frontiers in Education Conference, FIE*, 2003, vol. 1, p. T4G1-T4G8. doi: 10.1109/FIE.2003.1263392.

[8] R. S. Lindell, E. Peak, and T. M. Foster, “Are They All Created Equal? A Comparison of Different Concept Inventory Development Methodologies,” in *https://doi.org/10.1063/1.2508680*, 2007. doi: 1.2508680.

[9] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*, 7th ed. John Wiley & Sons, 2014.