PHYSICS EDUCATION RESEARCH SECTION

The Physics Education Research Section (PERS) publishes articles describing important results from the field of physics education research. Manuscripts should be submitted using the web-based system that can be accessed via the American Journal of Physics home page, http://ajp.dickinson.edu, and will be forwarded to the PERS editor for consideration.

Response times to conceptual questions

Nathaniel Lasry
Physics Department, John Abbott College, Montreal, Quebec, Canada H9X 3L9

Jessica Watkins[a)] and Eric Mazur
School of Engineering and Applied Sciences, Harvard University, 9 Oxford Street, Cambridge, Massachusetts 02138 and Department of Physics, Harvard University, 9 Oxford Street, Cambridge, Massachusetts 02138

Ahmed Ibrahim
Faculty of Education, McGill University, 3700 McTavish Street, Montreal, Quebec, Canada H3A 1Y2

(Received 6 December 2011; accepted 14 June 2013)

We measured the time taken by students to respond to individual Force Concept Inventory (FCI) questions. We examine response time differences between correct and incorrect answers, both before and after instruction. We also determine the relation between response time and expressed confidence. Our data reveal three results of interest. First, response times are longer for incorrect answers than for correct ones, indicating that distractors are not automatic choices. Second, response times increase after instruction for both correct and incorrect answers, supporting the notion that instruction changes students' approach to conceptual questions. Third, response times are inversely related to students' expressed confidence; the lower their confidence, the longer it takes to respond.
[http://dx.doi.org/10.1119/1.4812583]

I. INTRODUCTION

Physics education research lies at the crossroads of physics, education, and psychology and uses methods derived from all three disciplines. Connections between physics and psychology can be traced back at least to Hermann von Helmholtz, who, in addition to mentoring historical figures in physics including Albert Abraham Michelson, Max Planck, and Wilhelm Wien, also mentored Wilhelm Wundt, the scholar traditionally credited for the birth of psychology.[1] Helmholtz pioneered reaction-time research by measuring "the time that it took for a nerve impulse to pass through the severed leg of a frog. He then succeeded in adapting the method for human subjects, who were asked to respond by pushing a button whenever a stimulus was applied to their leg."[2] Since Helmholtz, subject reaction times continue to be used in cognitive psychology.[3]

Looking at longer time scales and higher cognitive levels, psychometricians have used response times on achievement tests to study student effort,[4,5] response strategies,[6] and speededness.[7] Response times have been related to ability, providing additional data for item-response theory and other scoring models.[8,9] However, research shows that the relationship between response times and accuracy is difficult to generalize because it is dependent on the content and context of the test.[10] Additional research is needed to link response times to cognitive and educational psychology research,[10] for better understanding of the cognitive processes in answering test questions.

In this short paper, we show data suggesting that response times can provide additional insight into students' cognition during conceptual assessments. We obtained response times using a computerized version of the Force Concept Inventory (FCI).[11,12] We find significant differences in response times between correct and incorrect answers. We also find that students take longer answering questions after instruction (post-test) than they do before instruction (pre-test). Finally, we show a negative relationship between response time and student confidence.

II. METHODS

We collected FCI data from students ($N = 894$) in introductory mechanics courses at Harvard University between 1991 and 1997. These introductory courses were taught using Peer Instruction.[13,14] At the beginning and end of the term, each student took a computerized version of the FCI, which displays only one question at a time. After selecting an answer and indicating their confidence in the answer, the next question is displayed. The students were not allowed to return to a previously answered question. Response times were computed as the difference in time stamps between the displays of consecutive questions.

III. RESULTS

Table I shows the average time taken for correct $\langle t_{corr} \rangle$ and incorrect $\langle t_{inc} \rangle$ answers on the FCI, both before and after instruction. Before instruction, the average correct answer

Table I. Average time taken (in seconds) for correct and incorrect answers on the FCI pre-test and post-test.

	$\langle t\rangle_{pre}$ (s)	$\langle t\rangle_{post}$ (s)	$\langle t\rangle_{post} - \langle t\rangle_{pre}$ (s)
Correct	35.14	38.84	3.69[a]
Incorrect	46.74	59.24	12.50[a]
Difference	11.60[a]	20.40[a]	

[a]$p < 0.001$.

response time, $\langle t_{corr}\rangle_{pre} = 35$ s, is significantly shorter ($p < 0.001$) than the average time taken to give incorrect answers, $\langle t_{inc}\rangle_{pre} = 47$ s. At the end of the semester, response times increase significantly ($p < 0.001$) for all answers ($\langle t_{corr}\rangle_{post} = 39$ s; $\langle t_{inc}\rangle_{post} = 59$ s), and incorrect answers take 50% longer than correct answers on the post-test. We used non-parametric tests to evaluate statistical significance because response times do not distribute normally, a finding consistent with the literature.[15]

Figure 1 shows a histogram of the response times for correct and incorrect answers for a typical question. Looking at individual FCI questions, we expect "longer" questions to have greater response times. Although questions with more words are longer, many FCI questions also contain pictures. Determining question length is non-trivial: the time needed to parse a picture cannot be compared easily to the time taken to read text. To control for question length, we examine student response times on each FCI question individually and compare the time taken to answer a given question before instruction with the time taken to answer the same question after instruction. If no association is expected between correctness and response time, then responding correctly should take more time for half of the questions and less time for the other half. We determine the number of questions for which $\langle t_{inc}\rangle$ is greater than $\langle t_{corr}\rangle$, both for the 1992 version (29-item) of the FCI and the 1995 version (30-item) and calculate the related binomial probabilities. On the 1992 version of the FCI, $\langle t_{inc}\rangle$ is greater than $\langle t_{corr}\rangle$ for 26 out of 29 items ($p < 0.0001$) before instruction and for 27 items ($p < 0.0001$) after instruction; on the 1995 version of the FCI, $\langle t_{inc}\rangle$ is greater than $\langle t_{corr}\rangle$ for 25 out of 30 items ($p = 0.0001$) before instruction and for 28 items ($p < 0.0001$) after instruction (Table II).[16]

Figure 2 shows the dependence of students' average response times for incorrect answers on their pretest score.

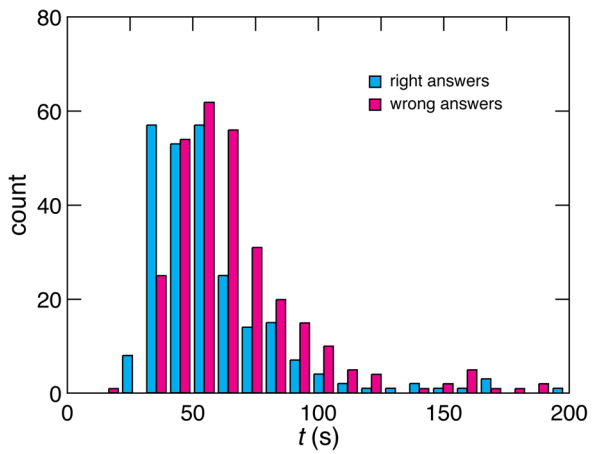

Fig. 1. Distribution of response times for incorrect and correct answers on a typical FCI question (question #13 on the 1992 version of the FCI).

Table II. Number of questions for which $\langle t_{corr}\rangle$is smaller/greater than $\langle t_{inc}\rangle$on the 29-item 1992-version and on the 30-item 1995-version of the FCI; both at the beginning (pre-test) and end (post-test) of the term.

	# questions with $\langle t_{corr}\rangle$ smaller than $\langle t_{inc}\rangle$	# questions with $\langle t_{corr}\rangle$ greater or equal to $\langle t_{inc}\rangle$	$\chi 2$
1992 FCI pretest	26	3	$p < 0.0001$
1992 FCI posttest	27	2	$p < 0.0001$
1995 FCI pretest	25	5	$p < 0.0001$
1995 FCI posttest	28	2	$p < 0.0001$

The data show a small increase in average time on incorrect answers as background knowledge increases. The correlation between pretest score and $\langle t_{inc}\rangle_{pre}$ is quite small (FCI-1992: $r = 0.035$; FCI-1995: $r = 0.053$).

Figure 3 displays the number of correct and incorrect answers for each confidence level (0 = just guessing, 1 = not quite sure, 2 = very sure). Few students rate their answers with low confidence and the majority of questions were answered correctly and with high confidence.

Figure 4 shows the average response times for each confidence level on the pretest, $\langle t\rangle_{pre}$. We find similar trends for posttest answers. Because of the large sample size, the standard errors are very small: 1.32 s, 0.43 s, and 0.18 s for confidence levels of 0, 1, and 2, respectively. Response times decrease significantly ($p < 0.05$) with each increment of expressed confidence. When students are guessing (confidence level 0), there is no significant difference in response time between correct and incorrect answers. However, we do observe significant differences at confidence levels 1 and 2.

IV. DISCUSSION

A. Incorrect answers take more time

When response times are used in perceptual-motor tasks, "speed-accuracy tradeoffs" become important. A speed-accuracy tradeoff describes the idea that the faster participants try to accomplish a perceptual-motor task, the likelier they are to make errors. With respect to conceptual questions, we find a reversed speed-accuracy relationship because correct answers are given faster than incorrect ones.

Conceptual surveys such as the FCI are built using common-sense beliefs that can act as strong distractors.[11,17]

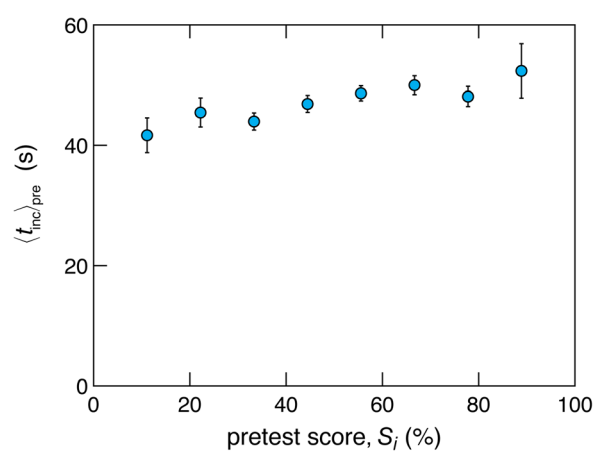

Fig. 2. Average response time for incorrect answers on the pre-test as a function of the pre-test score.

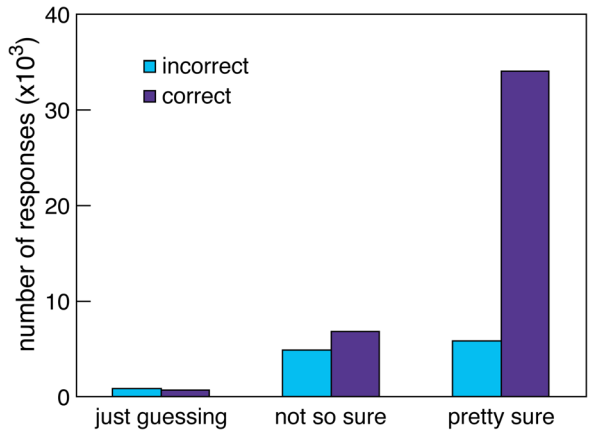

Fig. 3. Number of correct and incorrect answers on the pretest for each level of confidence.

Common-sense beliefs are often portrayed as "very stable, and conventional physics instruction does very little to change them."[17] These are ideas that "have been firmly held by some of the greatest intellectuals in the past, including Galileo and even Newton. Accordingly, these common-sense beliefs should be regarded as reasonable hypotheses grounded in everyday experience."[18] When the FCI was published, interviews with introductory physics students found that students usually have concrete explanations for their answers.[11]

If students' incorrect answers are common-sense beliefs derived from everyday experiences, then one would expect that students choose these distractors automatically, yielding short response times. However, our data show that incorrect answers systematically take *more* time than correct ones, indicating that the distractors are not automatic choices for our population. This result might be due to our sample selection, which is heavily biased toward high-background knowledge students. In Fig. 1 we report the dependence of incorrect response times on background knowledge. We only find a weak dependence of incorrect response time on background knowledge.

B. Posttest takes more time

The ability to perform better on a test that has previously been seen is often referred to as a "practice-effect." Our findings go against the idea of a practice effect because if students were familiar with the questions, one would expect shorter response times on the post-test, whereas we find longer response times. The data in Table I show that students take more time on FCI questions after instruction, even though they have seen the same test at the beginning of the term.

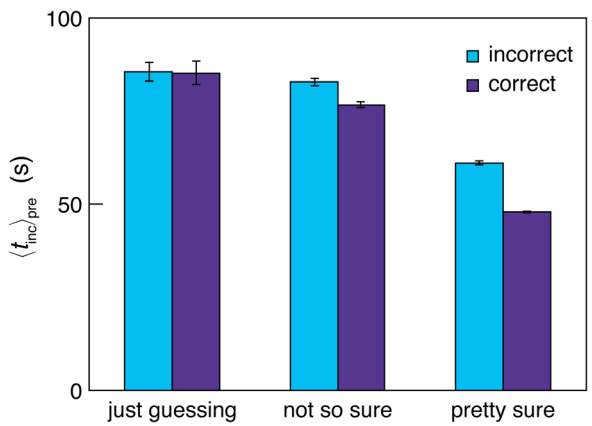

Fig. 4. Average response time for correct and incorrect answers at each level of student confidence. The error bars indicate one standard error.

We have also reported that correct answers have shorter response times. However, although there are more correct answers after instruction, we find on average that student response times increase at the end of the semester. One possibility for longer post-test response times could be how students "frame" the questions.[19,20] Before instruction, students may view the conceptual questions on the FCI as "simplistic" questions, not "real physics" questions. During a semester of Peer Instruction,[21] however, the students discover that they get many of the conceptual questions in class wrong, forcing them to rethink their understanding when answering conceptual questions. At the end of the semester, students may no longer see conceptual questions as simplistic but rather frame them as questions that may require application of physics concepts. Longer response times suggest that students spend more time thinking about the question after instruction.

C. Response time decreases with increased confidence

Figure 4 shows that confident students take less time to answer FCI questions, indicating they spend less time weighing alternatives. Interestingly, these students spend more time thinking about incorrectly answered questions than about correctly answered ones. This finding suggests that these incorrect answers that reflect common-sense beliefs are not automatic choices even when student express confidence in them.

Finally, the data in Fig. 3 show that most students in this study are highly confident in their answers. When students are just guessing, the number of questions answered correctly almost matches the number of questions answered incorrectly. This suggests that these students who state they are just guessing are likely to be choosing between two answers, one of which is correct. Therefore, students may not be randomly guessing between all five choices.

Our findings are limited by the selective nature of the student population in this study and, to some degree, the inclusion of the time taken to choose a confidence level within reported response times. To control for background knowledge, we correlated pretest score with response time, but did not find a significant correlation. To control for the time taken to rate confidence levels, we compared right and wrong answers at similar confidence levels and found that wrong answers take significantly more time except when students are guessing.

V. CONCLUSIONS

In this paper, we show that response times provide insight into students' thinking on conceptual surveys. For our student population, incorrect answers take more time than correct answers, showing that commonsense distractors do not trigger automatic responses from students. We also find that students take longer on post-test questions than on pre-test questions, which argues against the notion of a practice-effect on the FCI. Finally, we find that average response times decrease with increased confidence.

Taken together, our results suggest that response times can help assess how students are thinking about a concept. Quick incorrect responses suggest automatic alternate conceptions. Quick correct responses suggest a well-ingrained Newtonian concept. Slow responses suggest a hybrid conceptual state with neither correct nor incorrect automatic responses. Thus,

response times can augment the diagnostic power of the FCI and provide an additional interesting metric for research.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grant No. DUE-0716902. The authors thank D. van Wieren for assisting with the data analysis, Jason Dowd, Orad Reshef, and Kevin Vora for their help with the editing process, and Brian Frank for his input on an early version of the manuscript. N.L. would also like to thank Howard Gardner for helpful discussions.

[a)]Present address: Department of Education and Center for Engineering Education and Outreach, Tufts University, Medford, Massachusetts 02155

[1]Michael Cole, *Cultural Psychology: A Once and Future Discipline* (Harvard U.P., Cambridge, MA, 1996).
[2]Howard Gardner, *The Mind's New Science: A History of the Cognitive Revolution* (Basic Books, New York, NY, 1985).
[3]R. Ratcliff, T. van Zandt, and G. McKoon, "Connectionist and diffusion models of reaction time," Psychol. Rev. **106**(2), 261–300 (1999).
[4]S. L. Wise and C. E. DeMars, "An application of item response time: The effort-moderated IRT model," J. Educ. Meas. **43**(1), 19–38 (2006).
[5]S. L. Wise and X. J. Kong, "Response time effort: A new measure of examinee motivation in computer-based tests," Appl. Meas. Educ. **18**(2), 163–183 (2005).
[6]R. Mislevy, "Foundations of a new test theory," in *Test theory for a new generation of tests*, edited by N. Frederiksen, R. Mislevy, and I. Bejar (Lawrence Erlbaum Associates: Hillsdale, NJ, 1993).
[7]D. L. Schnipke and D. J. Scrams, "Modeling item response times with a two-state mixture model: A new method of measuring speededness," J. Educ. Meas. **34**(3), 213–232 (1997).
[8]F. Abdelfattah, "Response latency effects on classical and item response theory parameters using different scoring procedures," in *College of Education* (Ohio University, OH, 2007), p. 167.
[9]K. K. Tatsuoka and M. M. Tatsuoka. "A model for including response time data in scoring achievement tests," in *1979 computerized adaptive testing conference* (Minneapolis, MN., 1997).
[10]D. L. Schnipke and D. J. Scrams, "Exploring issues of examinee behavior: Insights gained from response time analyses," *in Computer-Based Testing: Building the Foundation for Future Assessments*, edited by C. N. Mills *et al*. (Lawrence Erlbaum Associates, Mahwah, NJ, 1997).
[11]D. Hestenes, M. Wells, and G. Swackhammer, "Force Concept Inventory," Phys. Teach. **30**, 141–158 (1992).
[12]I. A. Halloun *et al*., *Revised Force Concept Inventory*, appearing in Ref. 14, 1997.
[13]C. Crouch and E. Mazur, "Peer Instruction: Ten years of experience and results," Am. J. Phys. **69**(9), 970–977 (2001).
[14]E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
[15]W. van der Linden, "A hierarchical framework for modeling speed and accuracy on test items," Psychometrika **72**(3), 287–308 (2007).
[16]See supplementary material at http://dx.doi.org/10.1119/1.4812583 for more detailed information on response time per question.
[17]I. A. Halloun and D. Hestenes, "Common-Sense Concepts About Motion," Am. J. Phys. **53**(11), 1056–1065 (1985).
[18]I. A. Halloun and D. Hestenes, "The initial knowledge state of college physics. Students," Am. J. Phys. **53**(11), 1043–1055 (1985).
[19]D. Hammer, A. Elby, R. E. Scherr, and E. F. Redish, "Resources, framing, and transfer," in *Transfer of Learning: Research and Perspectives*, edited by J. Mestre (Information Age Publishing, Charlotte, NC, 2004).
[20]E. F. Redish, "A theoretical framework for physics education research: Modeling student thinking," in *Enrico Fermi Summer School, Course CLVI 2004* (Italian Physical Society, Varenna, Italy, 2004).
[21]E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, N.J., 1997).

Spitzer Telescope

This achromatic refractor with a six inch aperture came to Kenyon College in the 1950s from the astronomer Lyman Spitzer and his family. It is marked "Gall & Lemke", a New York optics firm founded in 1842. This instrument, which is lacking drives, was certainly intended for amateur use. When I came to Kenyon in 1964 Franklin Miller would set it up in the middle of the campus to give students the opportunity to view astronomical objects. It is now in the Greenslade Collection. There is another Spitzer Telescope orbiting the earth and taking data in the infrared wavelength region. Notes and photograph by Thomas B. Greenslade, Jr., Kenyon College)