

Effects on assessment caused by splits between belief and understanding

Timothy L. McCaskey, Melissa H. Dancy, and Andrew Elby

TM: University of Maryland, College Park MD 20742-4111, mccaskey@physics.umd.edu

MD: Western Carolina University, Cullowhee NC 28723, dancy@email.wcu.edu

AE: University of Maryland, College Park MD 20742-4111, elby@physics.umd.edu

We performed a new kind of FCI study to get at the differences between what students believe and what they think scientists believe. Students took the FCI in the standard way, and then made a second pass indicating “the answer they really believe” and “the answer they think a scientist would give.” Students split on a large number of the questions, with women splitting more often than men.

Introduction

Do the students taking standardized concept tests such as the Force Concept Inventory (FCI) [1] believe the responses they give, or do the answers they really believe disagree with the answers they think scientists would give? The present study gets at this question. Although we display some intriguing data here, this project will not be complete until in-depth interviews explore the reasons behind students’ answers on the multiple-choice task described below. Our purpose here is to make plausible the idea that interesting things happen when students are asked to discriminate between belief and understanding. Chinn’s talk [2] originally inspired this idea.

The FCI Task

Three populations of introductory physics students were studied. Two were at the University of Maryland (UMd), a large, public university. One was a group of first-semester students taking the FCI *before* a “reformed” course. The other was a second-semester class, most of whom took a traditional first-semester course. The third group had just completed a traditional course at Davidson College (DC), a small, highly selective school.

Our task consisted of two parts. For the first part, students took the FCI (as given in Redish [3]) in the standard way. The only major difference between the two locations was that the UMd test came with a written instruction that says, “Avoid guessing. Your answers should reflect what *you* personally think.” The DC version lacked that

instruction. When finished with the test, students handed in their answer sheets and received instructions for the second part:

We’d now like you to take the Force Concept Inventory a second time. But this time, the instructions are a little different. First, write your name on the top of the test just as you bubbled it in on your scantron sheet. For each test item:

- Please **circle** the answer **you really believe**.
- Please draw a **square** around the answer **you think scientists would give**.

Here’s an example of how to mark your test.

- (A) Answer I really believe.
- (B) Some random answer.
- (C) Answer I think scientists would give.
- (D) Another random answer.
- (E) Yet another random answer .

If the answer you really believe agrees with the answer you think scientists would give, draw a circle and square around that same answer.

The DC instructions were the same, except students wrote answers (first pass, “belief,” and “scientist answer”) by hand on a separate sheet of paper.

We had to be careful with the UMd data. Students often misinterpreted the instructions, by skipping entire items or marking either a square *or* a circle, but not both. We disregarded *all* of a student’s data if they left five or more blanks. (A blank is a missing answer on the first pass through the FCI, or a missing square or circle on the second task.)

Data

We scored students' first pass (1), their circled "beliefs" (B), and their squared "scientist answers" (SA). For each student, we also counted the number of discrepancies between each pair of answer sets, for example, the number of times the student's "belief" differed from her "scientist answer." We counted these discrepancies as a genuine *split* only when neither of the relevant answers was a blank. Finally, in cases where the student's "belief" differed from his "scientist answer," we checked which answer the student bubbled in on his first pass.

In the following tables, the FCI "scores" are class averages, out of 30 questions. A split between "1" and "B" is when a student's non-blank answer on the first pass differed from the non-blank answer she indicated that she really believes. 1-SA and B-SA splits are defined similarly, where SA refers to the answer they think a scientist would give.

The counts for the first semester UMd class (PHYS 121) are as follows:

	Males	Females
N	59	57
first pass score	13.7	8.6
"belief" score	12.7	7.8
"scientist" score	13.2	8.6
1-B splits	5.9	6.8
1-SA splits	7.2	10.3
B-SA splits	5.5	8.5
... with 1=B	2.8	5.4
... with 1=SA	1.6	1.9

Table 1: PHYS 121 results by gender

The second semester UMd class data (PHYS 122) are in Table 2 below. These two classes share a number of common features. The females systematically scored lower than the males on all passes through the FCI, although unlike the males, their "scientist answers" provided their best scores.

	Males	Females
N	20	30
first pass score	17.2	11.3
"belief" score	16.5	10.8
"scientist" score	16.4	11.8
1-B splits	3.6	3.9
1-SA splits	5.0	9.0
B-SA splits	3.9	8.0
... with 1=B	2.5	6.3
... with 1=SA	1.2	1.3

Table 2: PHYS 122 results by gender

Note also that the bottom two rows do not add up to the total number of B-SA splits because a student's first-pass answer sometimes differed from *both* second-pass answers (B and SA). This test-retest reliability issue warrants further research.

An interesting gender difference to note is in the number of splits. Women tended to split responses more often. Also, in cases where the students split between their "belief" and "scientist answer," women had bubbled in the "belief" a higher percentage of the time (63% vs. 51% for the PHYS 121 students). This could mean a number of things. Women may trust their beliefs more than the things they hear from professors. It's also possible that they better heeded the instruction to bubble in "what *you* personally think." A third possibility is that the men reconcile their beliefs with their perception of the scientists' answers more frequently. A fourth, more mundane possibility is that students who score lower on the FCI, both men and women, also split more, in which case women's higher split rates could be partially or fully accounted for by their lower scores. Future interviews, combined with statistical analyses of a larger data set, will help point toward which of these possibilities is correct.

At Davidson College (DC), the FCI task was given as a post-test after one semester of traditional mechanics instruction in a small class setting. The DC data are given in Table 3 below:

	Males	Females
N	11	9
first pass score	20.9	19.8
“belief” score	21.1	16.8
“scientist” score	21.3	20.6
1-B splits	2.9	5.375
1-SA splits	2.5	3.625
B-SA splits	1.6	5.1
... with 1=B	0.45	1.7
... with 1=SA	1	3.2

Table 3: DC results by gender

The majority of the male B-SA splits came from two students. The rest either had either no splits at all (5 students) or only one (4 students). This data backs up one of the major UMd trends, namely, that women split their answers more often than the men do. They also score best with their “scientist” answers.

The DC study gave a space for comments. Many students seemed confused about why they would ever separate their belief and scientist answers. Here is a sample response:

“Ok, I’m a little confused about the purpose of this. My answers the second go around won’t be different than my answers the first time. Even if my answers are wrong, I wrote them believing that scientists or anyone else for that matter would give the same answer.”

The UMd study did not allow written comments, but several students expressed similar concerns vocally to the test administrator.

A Specific Cluster: N3

Four of the items on the FCI require knowledge of Newton’s Third Law (N3). We looked specifically at this cluster because we felt N3 is a counterintuitive concept that might lead to a lot of splitting. The N3 questions are written such that students are unlikely square the correct answer unless they learned about the scientifically-accepted in class. For the UMd PHYS 121 data, 46 students got at least two of the four N3 questions right on the SA pass. Of these, only 16 (35%) of the student did not have splits between their “beliefs” and “scientist answers” on any of

those four questions. Among the 15 students that got all of the N3 items right for their SA, only 3 had no splits. This suggests that people who “get” scientific concepts like Newton’s Third Law right on a standardized test often don’t really believe them or at least have doubts about them.

Further questions

A lot of questions about this data remain, and future interviews should help to answer them. The biggest unresolved issues center around the meaning of the B-SA splits. Students who split a lot may genuinely believe in an absolute sense that physicists are wrong about their ideas. We think it’s more likely that the students view their “split” answers as different yet equally valid ways of thinking about the same ideas.

Another issue is reconciliation: To what extent do high-splitters and low-splitters think it is possible and valuable to resolve discrepancies between their own ideas and the scientifically-accepted answers?

The gender differences bring up some interesting questions, too. Why do women split more? When men and women split their answers, do they do it for the same reasons?

Also, what are the implications for teachers? What type of student has the best chance of both understanding the “scientific” perspective and successfully reconciling it with experience to achieve deeper learning? Are different pedagogical approaches required for a high achiever who disregards her real beliefs in favor of a scientist answer, or a more mediocre student who splits a lot and seems willing (though unable) to reconcile? Finally, given this research about splits, what kind of information does the FCI really provide a teacher?

As stated earlier, we cannot definitively answer these questions without interview data and perhaps additional split data from various populations. However, from this preliminary look, we can say that there’s potentially a lot of interesting “stuff” hiding beneath what at first looks like a simple FCI task.

Epilogue:

Interviews and statistical analyses conducted since this paper was written address many of the issues just raised. See our Winter 2004 AAPT talk at <http://www.physics.umd.edu/perg/talks/MiamiAAPT/MiamiMcCaskey.pdf> for further details.

References:

- [1] D. Hestenes, M. Wells, and G. Swackhamer, "Force Concept Inventory," *Phys. Teach.* **30**, 141-158 (1992)
- [2] C. Chinn, "Knowledge, Belief, and Understanding in Learning Science," 126th AAPT National Meeting (2003)
- [3] E. F. Redish, *Teaching Physics with the Physics Suite*, John Wiley and Sons, Inc. (2003)