

Validating the Force Concept Inventory with Sub-Questions: Preliminary Results of the Second Year Survey

Jun-ichiro Yasuda, Masa-aki Taniguchi

Abstract

We address the validity of the FCI, that is, whether respondents who answer FCI questions correctly have an actual understanding of the concepts of physics tested in the questions. We used sub-questions that test students on concepts believed to be required to answer the actual FCI questions. Our sample size comprises about five hundred respondents; we derive false positive ratios for pre-learners and post-learners, and evaluate the significant difference between them. Our analysis shows a significant difference at the 95 % confidence level for Q.6, Q.7, and Q.16, implying that it is possible for post-learners to answer three questions without understanding the concepts of physics tested in the questions; therefore, Q.6, Q.7 and Q.16 are invalid.

Key words: physics education research, force concept inventory, validation.

INTRODUCTION

Numerous types of diagnostic tools have been studied to examine how much students have learnt physics. The Force Concept Inventory (FCI) is one of the most important instruments for assessing students' understanding of the Newtonian conceptual framework (Hestenes, Wells & Swackhamer, 1992; Halloun & Hestenes, 1985a, 1985b; Huffman & Heller, 1995; Heller & Huffman, 1995). The FCI is a 30-item, five-choice survey that can be solved without the use of equations. Further, the distractors in the questions are constructed based on the naive conceptions about mechanics.

When conducting a survey using a diagnostic tool such as the FCI, it is first necessary to analyse its validity (Redish, 2003: p. 96). Validity refers to whether the instrument measures what it claims to measure. In the case of the FCI, we must investigate whether the FCI accurately assesses students' conceptual learning of Newtonian mechanics.

The FCI has previously been validated from various standpoints. Hestenes and colleagues evaluated the validity of the wording and diagrams in its questions (Hestenes, Wells & Swackhamer, 1992; Hestenes & Halloun, 1995), while Rebello and Zollman analysed the validity of the distractors in the questions by comparing students' responses to four FCI open-ended questions (Rebello & Zollman, 2004). Morris and colleagues also evaluated the validity of the distractors by analysing the item response curves (Morris et al., 2006, 2012), and Stewart and colleagues validated the contexts of the questions using a ten-question context-modified test (Stewart, Griffin & Stewart, 2007). Yasuda and colleagues interviewed students and found that some students were able to provide the correct answer to Q.6, Q.7 and Q.16 even when using the incorrect reasoning (Yasuda, Uematsu & Nitta, 2012).

In our approach, we use a decision table to clear the problem (Table 1). In Table 1, the rows mean whether a student answers an FCI question correctly or not, and the columns mean whether the student understands the concept tested in the FCI question or not. False positives refer to correct answers provided by students who do not understand the physics concept being tested in the questions [1]. False negatives, by contrast, refer to incorrect answers provided by students who understand the physics concept tested in the question. The FCI question may be valid if the true positives and true negatives are many enough, and the FCI question may not be valid if the false positives and false negatives are many enough.

Table 1: Decision Table of an FCI question

	Understanding	NOT Understanding
Correct	True positive	False positive
Incorrect	False negative	True negative

From Table 1, we tackle the following 3 issues:

1. **How can we define understanding?**

The definition of the word “understanding” is one of the difficult problems of the cognitive science. In our study, we will define understanding operationally by means of decomposed questions of the original FCI question.

2. **How can we evaluate the amount of false answers?**

There is a well-known statistical variable to quantify the amount of false answers. We will explain and use it later.

3. With the variables, how can we evaluate the validity?

Using the statistical variables, we need a criterion or a standard value to judge whether an FCI question is valid or not. In order to decide the criterion, we form a hypothesis on this issue later.

In simple terms, our research question examines whether students who respond correctly to an FCI question, understands the physics concept that a question is meant to test. We explain our methods in order of the three issues described above.

METHOD 1: DEFINITION OF UNDERSTANDING

Usually, students answer an FCI question and we check whether the answer is correct or not. However, we cannot judge if the student understands the concept tested in the question. Therefore, we decompose an FCI question into a series of cognitively sequenced questions (Figure 1). We refer to these questions as *sub-questions*. If a student answers all of the sub-questions correctly, we assume that he or she has an understanding of the physics concept tested. The decision table of answers with sub-questions is presented in Table 2.

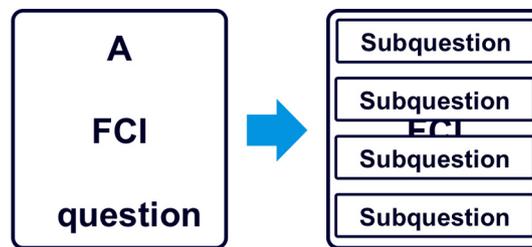


Figure 1: Decomposition of an FCI question

Table 2: Decision Table of an FCI question with sub-questions (SQs)

	Answer all SQs correctly	Answer not all SQs correctly
Correct	True positive	False positive
Incorrect	False negative	True negative

Which of the eight choices best represents the direction of the following variables, just after the string breaks? If you think a variable is zero, write 9.

- SQ1. Force acting on the ball
- SQ2. Acceleration of the ball
- SQ3. Velocity of the ball

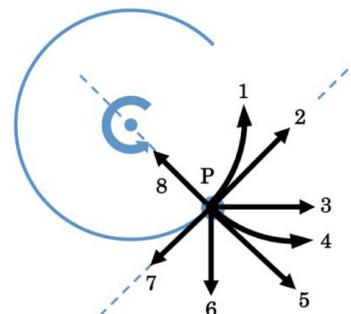


Figure 2: Outline of the sub-questions (SQs) of FCI Q.7

As an example, we show the outline of the sub-question of the FCI Q.7 in Figure 2. The original FCI Q.7 probes students to comment on the trajectory of the ball after the string breaks. The sub-questions presented in Figure 2 gives more direct information such as force, acceleration and the velocity of the ball after the string breaks [2].

METHOD 2: QUANTIFICATION OF FALSE POSITIVES

We analyze the false positives by evaluating a well-known statistical variable, *false positive ratio*. If event A represents answering an FCI question correctly and event B represents answering all the related sub-questions correctly, then the false positive ratio of that question is defined as follows:

$$P(A | \text{NOT } B) = \frac{N(A \text{ and NOT } B)}{N(\text{NOT } B)}$$

where $N(A \text{ and NOT } B)$ is the number of students who answered an FCI question correctly and answered more than one of the sub-questions incorrectly, and $N(\text{NOT } B)$ refers to the number of students who answered more than one of the sub-questions incorrectly. In this case, the false positive ratio can be interpreted as the identification of the subgroup that does not understand the physics concept and calculating the percentage of correctly answered questions (Figure 3).

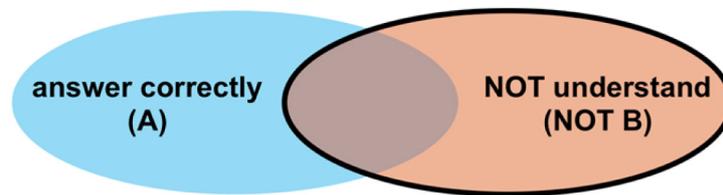


Figure 3: Venn diagram about false positive ratio

METHOD 3: CRITERION OF VALIDITY

We need a criterion, namely a reference value, to relate the false positive ratio to the validity. The reference value is the “ideal” probability with which a student who does not understand the concept tested answers correctly. If a false positive ratio of an FCI question is much larger than the reference value, the FCI question is judged to be invalid.

The simplest reference value is the probability to answer correctly by random guessing, that is, $1/5 = 0.2$. However, students who misunderstand the concept might tend to choose a wrong answer if the distractors of the question are well constructed, or these might tend to choose a right answer if the distractors of the question are not well constructed. In the former case, the ideal probability is less than 0.2, and in the latter case, the ideal probability is more than 0.2.

Since we need to separate the effect of distractors, we take, as the reference value for each question, the probability with which a student who *has not learnt* (pre-learner) the concept tested but answers correctly: FPR_{pre} [3]. This value is then compared with the probability with which a student who *has learnt* (post-learner) the concept tested and answers correctly: FPR_{post} . If the structure of the question is valid, it follows that only if students cannot understand the physics concept will they answer incorrectly, except in cases of coincidence. Therefore, if we choose the subgroups that do not understand the physics concept tested from both pre-learners and post-learners, the percentage of questions answered correctly for each subgroup should be comparable. However, there is one case in which FPR_{pre} and FPR_{post} are not comparable i.e. when the post-learner responds correctly by

using an incorrect physics concept or by remembering the correct answer of a similar question. In this case, the false positive ratio of post-learners could become large. Therefore, if FPR_{post} is significantly larger than FPR_{pre} , we judge that the question is invalid because post-learners can correctly answer the question even if they have no understanding of the physics concept tested.

We can explain this criterion from another standpoint. We begin with forming the following hypothesis: if an FCI question is valid, the FCI question cannot distinguish whether the student has already learnt the concept or not when a student does not understand the concept tested [4]. In this case, the false positive ratio of the pre-learners takes similar value to the false positive ratio of the post-learners. If we take the contraposition of this hypothesis, it follows that an FCI question is invalid if there is a significant difference between the value of the false positive ratio of pre-learners and the false positive ratio of the post-learners. The outline of this logic is shown in Figure 4.

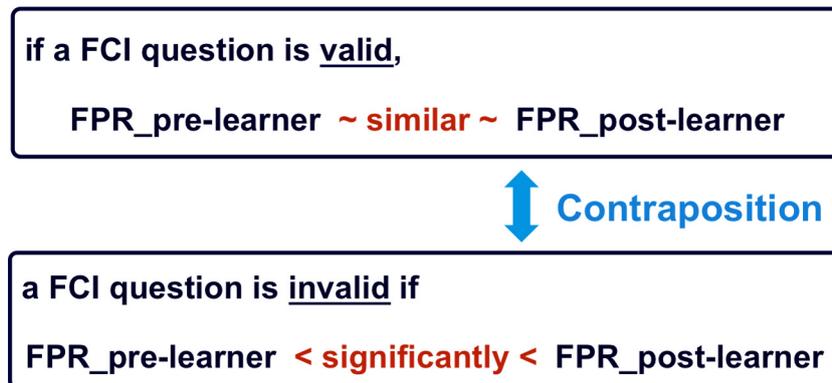


Figure 4: Outline of the logic about our criterion of validity

Table 3: False positive ratios of the pre- and post- learners in this and previous survey

	Q.5	Q.6	Q.7	Q.16
FPR_{pre} (2013)	0.11	0.576	0.35	0.13
FPR_{post} with CL95 % error (2013)	0.08 ± 0.06	0.653 ± 0.067	0.61 ± 0.07	0.52 ± 0.05
Significant difference	NO	YES	YES	YES
FPR_{pre} (2012)	0.39	0.26
FPR_{post} with CL95 % error (2012)	0.71 ± 0.13	0.50 ± 0.14
Significant difference	YES	YES

SETTINGS

DATA COLLECTION

We surveyed 524 students at one public university (Gifu U.) and three private universities (Meijo U., Kansai U. and Ritsumeikan U.) from April to June 2013. Respondents comprised students from different departments (e.g.: engineering, agriculture, human studies), and most were students in the university's physics classes (e.g., calculus based mechanics, general physics). The students were given no incentive to participate (in the form of money or grade points).

SURVEYED QUESTIONS

We surveyed the questions that showed false positives from our previous interview study (Yasuda, Uematsu & Nitta, 2012). The questions are Q.6, Q.7, and Q.16. For example, students were able to provide the correct answer to Q.16 even when using the incorrect reasoning that the forces were balanced because the two vehicles were moving at a constant speed. Similar shortcomings have been highlighted by other studies (Thornton et al., 2009; Scott, Schumayer & Gray, 2012). In addition to these questions, for comparison purposes, we surveyed Q.5, which showed no false positives in the interview. The physics concepts tested in each question are as follows; Q.5: circular motion, Q.6: circular motion, Q.7: circular motion, Q.16: Newton's third law (Hestenes & Jackson, 1992).

RESULTS

The results of our survey are presented in Table 3 which includes the errors of the false positive ratios of the post-learners (FPR_{post}) at the confidence level 95 %. If a false positive ratio of the pre-leaners (FPR_{pre}) is out of the error range, we judge that there is a significant difference between the FPR_{post} and the FPR_{pre} . With this criterion, we can see that there is a significant difference on Q.6, Q.7, and Q.16, and there is no significant difference on Q.5. Since Q.5 is the question for comparison, these results are consistent with our previous results. As for Q.6, the FPR_{pre} is just outside of the error bar, because the FPR_{pre} is considerably large. We think this is because the pre-learners can correctly answer Q.6 with knowledge from their daily experience.

We also show in Table 3 the results of our previous survey carried out in 2012 (Yasuda & Taniguchi, 2013). In this survey, we used similar sub-questions as for Q.7 but fewer respondents ($N = 111$). With the Table 3 and its plot, Figure 5, it is clear that these two surveys are consistent and the precision of the data is improved.

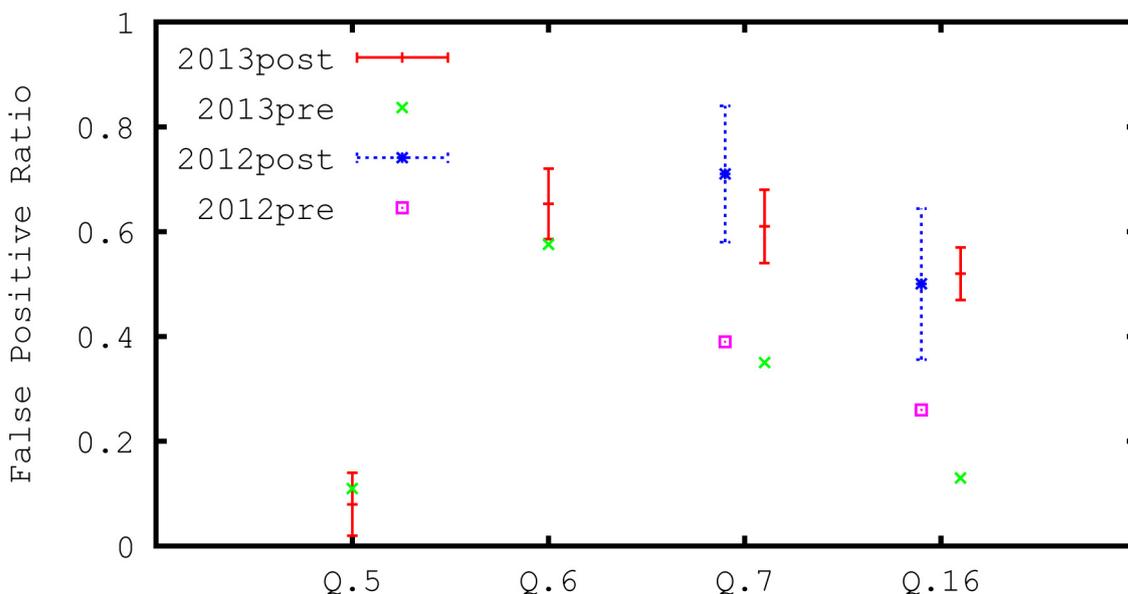


Figure 5: Plot of Table 3. The results of the post-learners are displayed with error bars at 95 % confidence level. The number of respondents is 524 for the survey in 2013 and is 111 for the survey in 2012

CONCLUSIONS

We evaluated the validity of the FCI using sub-questions and the false positive ratio. The false positive ratios of Q.6, Q.7 and Q.16 indicated that these questions are inadequate at the 95 % confidence level. This result implies that it is possible for post-learners to answer these questions without understanding the concepts of physics tested in the questions. This might be because the post-learners can correctly answer questions by using an incorrect physics concept or by remembering the correct answer of a similar question [5].

On the other hand, the false positive ratio of Q.5 indicated that Q.5 is a valid question and we have found no sign of the false positive on the other 26 questions from the interview study. Therefore, we can expect that 90 % of the FCI questions are adequate.

As part of future work, we need to confirm whether those 26 questions are adequate. Moreover, as for the generality, we need to confirm whether our results are true for the students in other countries. We also might need to confirm whether our results are changed if we use different types of sub-questions and evaluate the validity of the sub-questions. However, we should also think how far one should evaluate the sub-questions.

Further future work includes a plan to quantify the validity and estimate the systematic error of the total FCI score. The validation of the FCI has suggested the modification of the inadequate questions, but it might be difficult to compare the data of the modified FCI with the accumulated data. Instead, it will be better to evaluate the systematic error of the FCI from the evaluation of the validity. With this evaluation, we can continue to use the present FCI with reliable limitation.

REFERENCES

- Halloun, I. & Hestenes, D. (1985). The initial knowledge state of college physics students. *Am. J. Phys.*, 53(11), 1043–1055.
- Halloun, I. & Hestenes, D. (1985): Common sense concepts about motion. *Am. J. Phys.*, 53(11), 1056–1065.
- Hestenes, D. & Halloun, I. (1995). Interpreting the force concept inventory: A response to March 1995 critique by Huffman and Heller. *Phys. Teach.*, 33(8), 502–506.
- Heller, P. & Huffman, D. (1995). Interpreting the force concept inventory: A reply to Hestenes and Halloun. *Phys. Teach.*, 33(8), 503–511.
- Hestenes, D. & Jackson, J. (1992). Revised Table II for the Force Concept Inventory. Available at <http://modeling.asu.edu/R&E/Research.html>
- Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force concept inventory. *Phys. Teach.*, 30, 141–158.
- Huffman, D. & Heller, P. (1995). What does the force concept inventory actually measure? *Phys. Teach.*, 33(3), 138–143.
- Morris, G. A., Branum-Martin, L., Harshman, N., Baker, S. D., Mazur, E., Dutta, S., Mzoughi, T., & McCauley, V. (2006). Testing the test: Item response curves and test quality. *Am. J. Phys.*, 74(5), 449–453.

- Morris, G. A., Harshman, N., Branum-Martin, L., Mazur, E., Mzoughi, T., & Baker, S. D. (2012). An item response curves analysis of the Force Concept Inventory. *Am. J. Phys.*, 80(9), 825–831.
- Rebello, N. & Zollman, D. (2004). The effect of distractors on student performance on the force concept inventory. *Am. J. Phys.*, 72(1), 116–125.
- Redish, E. (2003). *Teaching Physics with the Physics Suite*. Hoboken: Wiley.
- Scott, T. F., Schumayer, D. & Gray, A. R. (2012). Exploratory factor analysis of a Force Concept Inventory data set. *Phys. Rev. ST Phys. Educ. Res.*, 8(2), 020105-1–020105-10.
- Stewart, J., Griffin, H., & Stewart, G. (2007). Context sensitivity in the force concept inventory. *Phys. Rev. ST Phys. Educ. Res.*, 3, 010102-1–010102-6.
- Taniguchi, M. & Yasuda, J. (2014). Quantitative Validation of Japanese Translation of Force Concept Inventory using Sub-Questions. *Journal of the Physics Education Society of Japan (in Japanese only)*, 62(4), 226–231.
- Thornton, R. K., Kuhl, D., Cummings, K. & Marx, J. (2009). Comparing the force and motion conceptual evaluation and the force concept inventory. *Phys. Rev. ST Phys. Educ. Res.*, 5(1), 010105-1–010105-8.
- Yasuda, J. & Taniguchi, M. (2013). Validating two questions in the Force Concept Inventory with subquestions. *Phys. Rev. ST Phys. Educ. Res.*, 9(1), 010113-1–010113-7.
- Yasuda, J., Uematsu, H., & Nitta, H. (2012). Validating a Japanese version of the Force Concept Inventory. *Lat. Am. J. Phys. Educ.*, 6(1), 89–94.

NOTES

- [1] The FCI has five multiple-choice answers per question; therefore, if a respondent chooses the correct answer randomly, false positives will appear in 20 % of his or her answers. Thus, we should focus on highly frequent false positives.
- [2] You can see the sub-question of Q.16 in Yasuda & Taniguchi (2013), and you can get the original FCI questions in the following website of the American Modeling Teachers Association, <http://modelinginstruction.org/researchers/evaluation-instruments/fci-and-mbt/>
- [3] From the questionnaire responses, we determined whether students had studied the concept (for example, Newton's third law or uniform circular motion etc.) tested in FCI questions. If students answered yes to having studied either concept, we called them post-learners and pre-learners otherwise.
- [4] If we compare this hypothesis to the quantum states, this hypothesis is equal to the statement that the state: $| \text{not understanding} | \text{pre-learner} \rangle$ is degenerate with the state: $| \text{not understanding} | \text{post-learner} \rangle$ under the operator: FCI.
- [5] You can see the following study in Taniguchi & Yasuda (2014), in which we sufficiently developed the current analysis and added new results.

JUN-ICHIRO YASUDA

Center for General Education, Gifu University, Gifu, Japan

MASA-AKI TANIGUCHI

Comprehensive Scientific Education Center, Meijo University, Nagoya, Japan