Validating a japanese version of the force concept inventory

Jun-ichiro Yasuda[1], Haruko Uematsu[2], Hideo Nitta[2]
[1]*Department of Science and Technology, Meijo University, Nagoya, Aichi 468-8502, Japan.*
[2]*Department of Education, Tokyo Gakugei University, Koganei, Tokyo 184-8501, Japan.*

E-mail: jyasuda@meijo-u.ac.jp

(Received 16 September 2011; accepted 12 November 2011)

Abstract

We evaluated the Japanese version of the FCI that was translated by a group consisting of members from Tokyo University of Science and Tokyo Gakugei University. We interviewed 16 students using a semi-structured interview where they were asked to talk out loud as they took the survey. Using set criteria, we found false-positives, and we found false-negatives even with careful students. We also found that the decision of some students to choose the right answer was the result of guessing that came about from the inadequacy of the translation. We set criteria, one of which is based on the false-negatives, false-positives, and guessing that came about from the inadequacy of the translation. Under the set criteria, we found that, of the 30 questions, 16 questions had at least one of these problems.

Keywords: Validation, Force Concept Inventory (FCI), Diagnostic Tool, Concept Surveys.

Resumen

Hemos evaluado la versión japonesa del FCI, que fue traducido por un grupo formado por miembros de la Universidad de Tokio de Ciencia y la Universidad de Tokio Gakugei. Se entrevistaron a 16 estudiantes usando una entrevista semi-estructurada, donde se les pidió hablar en voz alta a medida que respondieron a la encuesta. Utilizando los criterios establecidos, encontramos los falsos-positivos, y nos encontramos con resultados falsos-negativos, incluso con estudiantes cuidadosos. También se encontró que la decisión de algunos estudiantes a elegir la respuesta correcta fue el resultado de adivinar que surgió de la insuficiencia de la traducción. Hemos establecido los criterios, uno de los cuales se basa en los falsos-negativos, falsos-positivos, y adivinando lo que surgió de la insuficiencia de la traducción. Según los criterios establecidos, se encontró que, de las 30 preguntas, 16 preguntas tenían por lo menos uno de estos problemas.

Palabras clave: Validación, Inventario del Concepto de Fuerza (FCI), Herramienta de Diagnóstico, Estudios de Concepto.

PACS: 01.40.Fk, 01.40.gf, 01.40.Ha. **ISSN 1870-9095**

I. INTRODUCTION

A variety of diagnostic tools have been advanced in the domain of physics to measure how student learning is improved. Among the diagnostic tools, the Force Concept Inventory (FCI) developed by Hestenes and his collaborators is a typical instrument used to assess student understanding of the Newtonian conceptual framework [1, 2, 3, 4]. The FCI has been translated into various languages and is widely used internationally in the field of physics education research. In Japan, the FCI is translated in Japanese by a group consisting of members from Tokyo University of Science and Tokyo Gakugei University [5] and also by Iida and Ishimoto [6] and the physics education research with the FCI is widely spreading [7].

The FCI is a 30-item 5-choice survey, which can be solved almost without equations. The distractors of the questions are constructed based on the knowledge of naïve conception of mechanics, which is found through the interviews to the students. Hestenes and Halloun suggest the criterion of the FCI score that the students with the FCI score above 85% reach Newtonian Mastery threshold, the students with the FCI score above 60% below 85% reach Newtonian entry threshold and have begun to use Newtonian concept coherently in their reasoning [8].

In General, when we do a quantitative survey with a diagnostic tool such as the FCI, we need to evaluate the *validity* of the diagnostic tool [9]. The validity is whether the instrument measures the construct it purports to measure [10]. The purpose of the FCI is to probe the student conceptual learning in Newtonian dynamics and the method is the questionnaire with multiple-choices. Therefore, in order to evaluate the validity of the FCI, we need to examine whether the questions of the FCI are constructed to probe the concepts of Newtonian dynamics of students accurately.

The FCI has been validated from various points of views. For example, Hestenes and his collaborators evaluated validity of the wording or the diagrams of the questions, that is, *the validity of the representation* of the questions and they explained it is beyond reasonable doubt [1, 8, 11]. Rebello and Zollman evaluated the validity of the distractors of the questions, by comparing students' responses on four FCI questions with similar responses to equivalent open-ended questions [12]. Stewart and his collaborators evaluated the validity of the contexts of the questions with a 10-question context-modified test [13].

In general, it is necessary to validate a survey translation, because the validity might vary depending on the group of the students [14]. Especially, in the case of the FCI where everyday speeches are used, the representations are not necessarily to be valid for Japanese students because of the cultural and educational difference between the United States and Japan. However, the validity of the Japanese versions of the FCI has hardly been evaluated. Therefore, in this work, we aim to evaluate the validity of the Japanese version of the FCI, translated by the group of Tokyo University of Science and Tokyo Gakugei University.

Among the several validities we mentioned above, the causes of the problems about the validity of the distractors and the contexts are attributed to the original FCI. On the other hand, the subjectivity and the interpretation by the translators affect the validity of the representation. Therefore, in this work, we aim to evaluate the validity of the representation.

We evaluate the validity of the representation of the Japanese version of the FCI with the method of semi-structured interview. This is because it is necessary to examine the thinking process of the students by asking flexible questions in addition to a series of prepared questions in order to see whether the students understand the intent of the questions accurately.

II. METHOD

A. Overview

In order to evaluate the validity of the representation of the Japanese version of the FCI, we interviewed total 16/students of Meijo U., Nagoya U. and Tokyo Gakugei U. with the method of semi-structured interview from the middle of July until the middle of September in 2010. Although we interviewed in most of the case, we asked graduate students to interview in part. We made the condition for the examinees to have learned the mechanics in high school or in university. Although some of the examinees took our classes, we didn't give them any scores to their grades, but gave them rewards. We interviewed the students one by one in the classroom or the seminar room in our university for one or two hours. We recorded the interview with an IC recorder or a video camera with agreement of the interviewee to preserve the statement of the interviewee accurately.

B. Procedure

At first, we asked the examinees to answer the Japanese version of the FCI within 30 minutes. We instructed them to read carefully the statements of the questions in order to decrease careless mistakes. Then, in the interview, we asked the examinees the reason why they chose their answers to each question. Although we prepared only this question, we asked flexible questions if the interviewer judged the explanations of the examinees were not enough. We instructed the examinees to talk out loud in order to clarify their thinking process. This method is based on the cognitive interview [10]. We didn't give the examinees the right answer during the interview, since it might be a hint of the following questions.

We sufficiently considered the ethic of the research. For example, we explained to the examinees the purpose of the research, the range of disclosure and the treatment of the personal information before the research. Then, we confirmed the agreement of the examinees.

C. Method of the Analysis

In order to evaluate the validity of the representation, we examined the number of *false-negatives* and *false-positives* among the answers of examinees [8, 15]. The false-negatives are the wrong answers that the examinees chose with the logic based on the Newtonian mechanics. The false-negatives include the careless mistakes. The false-positives are the right answers that the examinees chose without the logic based on the Newtonian mechanics. Since the FCI has 5-choices, the false-positive appears in 20% even if the examinees choose randomly.

We focus on the *false-negatives except careless mistakes* and the *false-positives with high frequency*, because their causes are attributed to the inadequacies of the questions rather than the incidental factors. Since the false-negatives except careless mistakes and false-positives with high frequency affect the FCI scores directly, we think they are so problematic. Therefore, in a question, even if just one examinee gives the false-negative except careless mistakes, we presume that the representation of the question is inadequate. And also, in a question, if the examinees give the false-positives with high frequency, we presume that the representation of the question is inadequate.

Moreover, we focus on *the right answers with guessing caused by the inadequacy of the question* (From here, we call this "the right answers with guessing"). The examinees who gave the right answer with guessing answer in the following process. At first, they explain with the logic based on Newtonian mechanics. If students are in the state of indecision caused by the inadequacy of a question, they guess by compensating their own condition and then give the right answer [16]. Since guessing is not based on the clear logic, the examinee who has guessed possibly changes their answers by the moment. And also, if an examinee gave the right answer with guessing in a question, the question has a risk that the other examinee with Newtonian

TABLE I. The sum and the corresponding numbers of the questions of the investigation items. There are 4 duplications in the numbers of the questions where there appeared the false-negatives except careless mistakes and the right answers with guessing. "(2)" on the right side of the numbers of questions indicates that there are two inadequacies in the same question.

Investigation Items	Sum of the questions (30 questions total)	Numbers of the questions
False-negative except careless mistakes	8 questions	1, 14, 18, 20, 21, 26, 27, 29
False-positive with high frequency	1 question	16
Right answer with guessing	11 questions (13 items)	1, 2, 3(2), 5, 8, 18, 19, 21(2), 22, 23, 27

concepts might give the wrong answer with similar guessing in turn, that is, give a false-negative. Therefore, in a question, even if just one examinee gives the right answer with guessing, we presume that the representation of the question is inadequate. We think the series of above presumptions are the most cautious one within our method.

III. RESULT

We show the result of the interview in Table I. We found 8 questions where there appeared a false-negative except careless mistakes, 1 question where there appeared a false-positive with high frequency, and 11 questions (13 items) where there appeared a right answer with guessing. There are 4 duplications in the numbers of the questions where there appeared the false-negatives except careless mistakes and the right answers with guessing. Therefore, it is 16 questions total which have inadequacies in their representations according to the criterion we presumed in this study.

We can classify the inadequacies we found in Japanese version of the FCI into 5 categories.

(1) Descriptions of physical situations that are ambiguous:

We found the examinees who showed indecision because in several questions there is no description about the existence of the air resistance or the friction or about the position of the observer. As an example, we take the Q.18 (see Fig. 1), which asks what forces are acting on the boy when he is at position P. Since there is no description of the observer, one may think several types of the forces, which are observed by the observer on the ground, on the swing and so on.

(2) Figures that cause students misunderstanding:

As an example, we take the Q.14 (see Fig. 2), which asks which of the paths would the bowling ball most closely follow after leaving the airplane. We found the examinees who showed indecision because there is no description of the position of the plane at the time when the bowling ball landed.

(3) Adverbs used to represent time that are ambiguous:

We found the differences among examinees to interpret the words which represent qualitatively the interval of the time, e.g. "immediately" (in Japanese, *"Suguni"*). As an example, we take the Q.27, which asks "If the woman suddenly stops applying the horizontal force to the block, the block will…." There are two choices "immediately comes to a stop" and "slowing to a stop". Here, it is no wonder that there is a student who interprets "immediately" represents one second and the time "slowing to a stop" represents a few of 10 seconds.

(4) Words that are unfamiliar to students:

As an example, we take the Q.21-Q.24, which asks the motion of the rockets in outer space. We found the examinees who don't know the meaning of the "thrust" or the physical situation in outer space (Actually, more detail explanation is described in the original FCI than in the Japanese version of the FCI).

(5) Structure of the question which tends to induce the false-positive with high frequency:

We found a question whose structure induces the false-positive with high frequency. For example, concerning the Q.16 (see Fig. 3), the right answer is 1 and the reason is the Newton's Third Law. However, among the 10 persons who gave the right answer, we found that 8 persons (80%) chose the right answer with the wrong reason that the forces are balanced because the two vehicles move in a constant speed. Since random choices have a 20% chance of false-positives, it is natural to think that this frequency (80%) is extraordinarily high.

In terms of the above classification of the inadequacies, we can summarize our research as Table II.

IV. DISCUSSION

1. Structural inadequacy of the questions:

The inadequacy of the Q.16 is serious because the inadequacy is not of the representation, but of the structure. That is to say, to the Q.16, the students can give the right

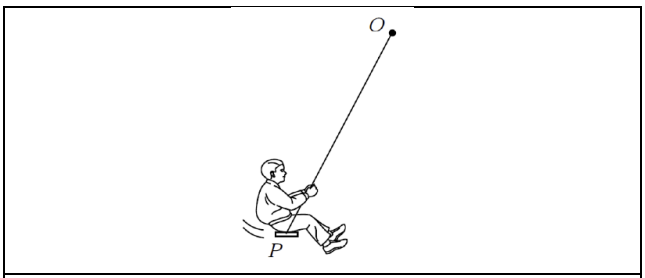

FIGURE 1. The figure of the question of swinging (Q.18).

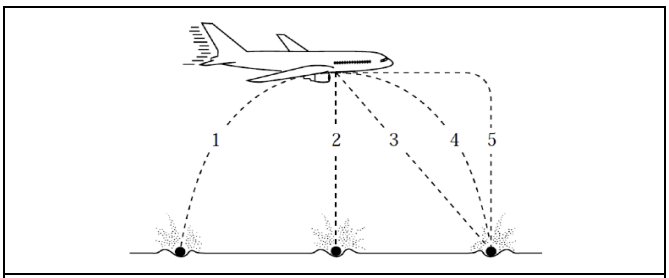

FIGURE 2. The figure of the question of plane (Q.14).

TABLE II. Summary of the evaluation of validity. We show the numbers of the examinees in the question where we found the false-negatives except careless mistakes ("FN" in the table) and the false-positives with high frequency ("FP" in the table).

Categories of the inadequacies	Numbers of Questions and the corresponding inadequacies
(1) Descriptions of physical situations that are ambiguous	Q.1, Q.2, Q.3 (whether there is the existence of air resistance), Q.1 (how early the ball reach the ground, FN:1), Q.3 (how far the force of the gravity is constant), Q.5 (position of the observer), Q.8 (ambiguity of the fraction of the magnitude of V_0 and V_k), Q.19 (whether "speed" means the instantaneous speed or the average speed), Q.20 (possibility of the inconstant force acting to the object, FN:1), Q.26 (whether there is the existence of air resistance, FN:1), Q.27 (whether there is the existence of the friction), Q.29 (whether the downward force by the air pressure means the net force or the component force, FN:1)
(2) Figures that cause student misunderstanding	Q.14 (position of the plane when the ball lands, FN:1), Q.18 (position of the point of the application of the force), Q.21 (the ambiguity of the difference of the trajectories of the rocket, FN:1)
(3) Adverbs used to represent a period that are ambiguous	Q.27 (how long the "immediately" represents, FN:1.)
(4) Words that are unfamiliar to students	Q.21, 22 (don't know the situation in outer space), Q.23 (don't know the term "thrust")
(5) Structure of the question which tends to induce the false-positive with high frequency	Q.16 (choosing the right answer with the wrong reason that the forces are balanced, FP:8)

answer with the wrong reason that the forces are balanced since the car and truck move in a constant speed rather than the right reason that there is Newton's Third Law.

In order to see whether the students understand the Newtonian concept probed in the Q.16, that is, third law for continuous forces [1], we examined the tendency of the answers to the Q.15, which probes the same concept as the Q.16 (The diagram and the choices of Q.15 is the same as that of the Q.16 (see Fig. 3.) and the question statement is changed from "After the car reaches the constant cruising speed at which its driver wishes to push the truck" to "While the car, still pushing the truck, is speeding up to cruising speed"). As the result, we found that the 5 out of 8 students who gave the right answer with wrong reason in the Q.16 gave the wrong answer to the Q.15. Therefore, it is possible that the 5 out of 10 (50%) students who gave the right answer to in the Q.16 don't understand the Newton's Third Law for continuous forces.

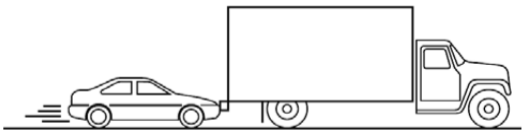

FIGURE 3. The figure of the Q.15 and Q.16.

A large truck breaks down out on the road and receives a push back into town by a small compact car as shown in the figure.
After the car reaches the constant crusing speed at which its driver wishes to push the truck,

1. the amount of force with which the car pushes on the truck is equal to that with which the truck pushes back on the car
2. the amount of force with which the car pushes on the truck is smaller than that with which the truck pushes back on the car
3. the amount of force with which the car pushes on the truck is greater than that with which the truck pushes back on the car
4. the car's engine is running so the car pushes againt the truck, but the truck's engine is not running so the truck cannot push back against the car. The truck is pushed forward simply because it is in the way of the car.
5. neither the car nor the truck exerts any forces on the other. The truck is pushed forward simply because it is in the way of the car.

2. Bias of the FCI scores in the students group:
We show the histogram of the FCI scores of the students group in Fig. 4. The average score is 70%, the maximum score is 93%, and the minimum score is 20%. According to the criterion by Hestenes and Halloun [8], the 6 students whose scores above 85% reach Newtonian Mastery threshold, the 5 students whose scores above 60% below 85% reach Newtonian entry threshold, and the 5 students whose scores below 60% score doesn't reach the threshold. From this fact, the bias of the FCI scores in the students group is thought to be little enough.

3. Certainty of our results:
The number of the examinees of our work, 16 persons, is not inferior to the number of examinees of previous works, e.g. [1]. The more the number of examinees is, the more the number of the questions where an examinee gave the false-negatives except careless mistake or the right answer with guessing more than just once will be. In this meaning, our result is still modest under our method.

The candidates of the questions where there might

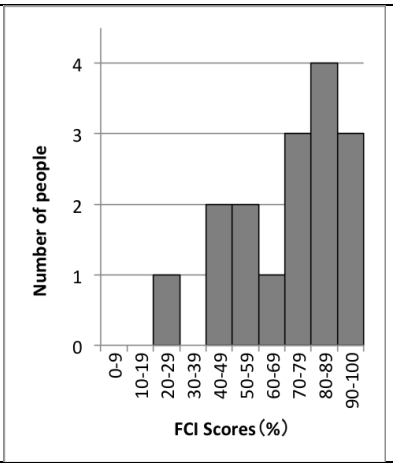

FIGURE 4. The histogram of the FCI scores of the students group.

appear the false-positives with high frequency are the Q.6 and Q.7, which ask "which of the paths would the ball most closely after…." In these questions, total 3 students gave the right answer with the wrong reason that the object is subjected to the force whose direction is tangential to the circumference. Although we have not consider these questions as the false-positives with high frequency since the frequency is not high enough, the frequency could be increased in the survey in the future.

4. Necessity to explain the physical situations in a question statement:

It could be thought that students should know naturally the situations about the friction or about the air resistance on the earth or in space, which we showed in Table II (1) or (4). However, the purpose of the FCI is not to probe the knowledge about the physical conditions on the earth nor in space but to probe the Newtonian concepts of the students [1]. Therefore, we supposed that we should give the students the information not directly concerned to probe the Newtonian concepts if at all possible.

On the other hand, in the FCI, Hestenes and his collaborators dare to use semi-realistic situations and everyday speech in order to set the context to be the student's personal resources for how the world works or the common sense belief [1] rather than what one is supposed to say in a physics class [17]. Therefore, it could be contrary to the purpose of the FCI to mention the air resistance or the friction in the question statements beyond the necessity.

5. The method of judging in the interview:

In the interview, we can clearly judge that a student *does not have* the Newtonian concepts or is guessing, since it is only necessary to pick up the student's statement concerning guessing just once. However, it is difficult to judge that a student *does have* the Newtonian concepts. This is because that it is necessary to judge the student's statement comprehensively, for example, whether there is a gap in their reasoning. In this meaning, the judgment of the false-negatives is more difficult than the judgment of the false-positives and the guessing.

6. The range and the prospect:

We have not evaluated the validity statistically, for example, how much error is induced by the inadequacies which we found, to the previous researches with the Japanese version of the FCI. This is because the number of interviewees of our work, 16 persons, is not enough to give a statistical conclusion with conviction. Therefore, our result will not affect the results of the previous researches immediately and it is necessary to evaluate the inadequacies statistically in the future work.

As a future work, we also plan to investigate whether the inadequacies of the Japanese version of the FCI might be also the inadequacies of the original FCI. In that study, we should take into account the cultural difference such that there could be the representations that frequently induce the false-negatives for Japanese students, but hardly induce the false-negatives for American students.

V. CONCLUSIONS

We evaluated the Japanese version of the FCI that was translated by a group consisting of members from Tokyo University of Science and Tokyo Gakugei University. We interviewed 16 students using a semi-structured interview and where they were asked to talk out loud as they took the survey. We set criteria, one of which is based on the false-negatives, false-positives, and guessing that came about from the inadequacy of the translation. Using the criteria, we found that, of the 30 questions of the translated FCI, 16 questions had problems. The effects of our result are limited, because we have not evaluated the validity quantitatively. On the other hand, we should correct the inadequacies caused by the mistranslation [18] as possible as possible. And also we should examine whether the other inadequacies correspond to the original FCI in the future.

We expect the knowledge we found will be utilized as the points to be considered not only when we modify the Japanese version of the FCI, but also when international researchers translate a diagnostic tool to any languages and when they develop a new diagnostic tool. And also, since only few researches of physics concepts by the interview has been done in Japan, we also expect the procedure of our work to be a model for the research from now on.

ACKNOWLEDGEMENTS

We have had the support of Prof. Taniguchi (Meijo U.) and have greatly benefited from the discussion with M. Hull (U. of Maryland). This work is supported in part by KAKENHI 22500801.

REFERENCES

[1] Hestenes, D., Wells, M. and Swachkamer, G., *Force Concept Inventory*, Phys. Teach. **30**, 141-158 (1992).

[2] Halloun, I. and Hestenes, D., *The initial knowledge state of the college physics students*, Am. J. Phys. **53**, 1043-1055 (1985).

[3] Halloun, I. and Hestenes, D., *Common sense concepts about motion*, Am. J. Phys. **53**, 1056-1065 (1985).

[4] Huffman, D. and Heller, P., *Interpreting the FCI: A replay*, Phys. Teach. **33**, 138-143 (1995).

[5] Kurihara, M., Tsukamoto, K., Funahashi, H., Negishi, A., Nitta, H., Takahashi, H., Kano, M., *Rikigaku Gainen Chosa (FCI) wayaku-ban no Sakusei to Jissi (in Japanese)*, Proceedings of the annual meeting of the Physics Education Society of Japan **26**, 38-39 (2009).

[6] The Japanese FCI translated by Iida and Ishimoto can be available in <http://modeling.la.asu.edu/R&E/Research.html>, visited in June 6 (2011).

[7] Tsukamoto, K., Higuchi, Y. and Kano, M., *University Students' Understanding of Force Concepts (in Japanese)*, Butsuri (in Japanese) **60**, 294-297 (2005).

[8] Hestenes, D. and Halloun, I., *Interpreting the force concept inventory: A response to March 1995 critique by Huffman and Heller*, Phys. Teach. **33**, 504-506 (1995).
[9] Redish, E., *Teaching Physics with the Physics Suite*, (Wiley, New York, 2003), p. 96.
[10] Desimone, L. M., Le Floch, K. C., *Are We Asking the Right Questions? Using Cognitive Interviews to Improve Surveys in Education Research*, Educational Evaluation and Policy Analysis **26**, 1-22 (2004).
[11] The *face validity* defined in [8] corresponds to the *validity of the representation* defined by us.
[12] Rebello, N. and Zollman, D., *The effect of distracters on student performance on the Force Concept Inventory*, Am. J. Phys. **72**, 116-125 (2004).
[13] Stewart, J., Griffin, H. and Stewart, G., *Context sensitivity in the force concept inventory*, Phys. Rev. ST Phys. Educ. Res. **3**, 010102 (2007).
[14] Yanai, H., Ogata, H., *SPSS data ni yoru toukei data kaiseki (in Japanese),* (Gendai-Sugaku Sha, Kyoto, 2006), pp. 169-170.
[15] Our definition of false-negative does not correspond to the definition in [8] exactly.
[16] If an examinee explains with the logic based on Newtonian mechanics and gives a wrong answer with guessing, the answer is presumed to be a false-negative.
[17] Redish, E., *Teaching Physics with the Physics Suite*, (Wiley, New York, 2003), p. 99.
[18] For example, the inadequacy such a Q.1 (how early the ball reach the ground, FN: 1) in (1) of the Table II is caused by the mistranslation of “considerably”.