

Isomorphic Test Of Newton's Third Law For Investigating Students' Scientific And Representational Consistency

Jusman Mansyur¹, Darsikin², and Syarif Hidayat²

^{1,2}Physics Education Department, Tadulako University

³SMPN 1 Dampelas, Kab. Donggala

E-mail: jusmansyurfis@yahoo.com¹, darsikinfis@gmail.com², syarifhidayat@yahoo.com³

ABSTRACT

We have developed a set of isomorphic test for investigating scientific and representational consistency in the context of Newton's Third Law. The test consisted of 30 multiple-choice items concerning five central force contexts: gravitation, electrostatics, magnetic, pushing, and crashing (impulse force). The test items were designed using various representations (i.e., verbal, diagram/vectorial and graphical). Before we conducted try out, test draft was reviewed by two physics content and evaluation experts for knowing appropriateness of concepts and isomorphic aspect of the test. We provide some evidence for analyses of the test based on the classical test theory. The limitation of the test is presented in this paper.

Keywords: isomorphic test, Newton's third law, scientific consistency, representational consistency

I. Introduction

Recent papers documented that student problem-solving competence varies (often strongly) with representational format, and that there are significant differences between the effects that traditional and reform-based instructional environments have on these competences (Kohl and Finkelstein, 2006). In literature on mathematics and physics education, a lot of attention is paid to student competence with different representational formats. By "representational format," we refer to the many different forms in which a particular concept or problem can be expressed and communicated, such as a graph, picture, free-body diagram, formula, etc. There is no purely abstract understanding of a physics concept—it is always represented in some form of representations. Therefore, being skilled in interpreting and using different representations and in coordinating multiple representations is highly valued in physics, both as a tool for understanding concepts and as a means to facilitate problem solving (Cock, 2012).

The role of multiple representations in learning is an important topic in the field of educational research. Multiple representations are often required for the understanding of scientific concepts and for problem solving. By "representational skills" we refer to students' ability to appropriately interpret and apply various representations of physics concepts and problems.

These different representations can include verbal, mathematical, graphical, and pictorial formats, though these categories are by no means

comprehensive or orthogonal (Kohl and Finkelstein, 2006). Nieminen *et al.* (2010) have investigated students' ability to interpret multiple representations consistently (i.e., representational consistency) in the context of the force concept. For the purpose, they have developed the Representational Variant of the Force Concept Inventory (R-FCI), which makes use of nine items from the 1995 version of the Force Concept Inventory (FCI). These original FCI items were redesigned using various representations (such as motion map, vectorial and graphical), yielding 27 multiple-choice items concerning four central concepts underpinning the force concept: Newton's first, second, and third laws, and gravitation.

Based on Nieminen *et al.* (2010), we designed 30 multiple-choice items test concerning five central concepts (i.e., gravitation, electrostatics, magnetic, pushing and crashing) by focusing on Newton's third law. For the purpose, we adopted a part of Nieminen *et al.* (2010) especially for Newton's third law and Bao *et al.* (2002).

We focused to the Newton's third law because of physics education research found that this concept is sensitive with contexts. The test could be used to identify the effect of instructional and reference books. Examples of Newton's third law in physics textbooks are dominated by context of continuous force (see Halliday and Resnick, 1994; Tipler, 1991). For instant, a student attracts a bonded rope; a book is on the table, etc. On the overall examples, authors engage the reader to consider the magnitude and orientation of the forces pair. From the example, author gives explanation of the role of Newton's third law. We think that teaching of Newton's third law in the classroom could be influenced by the widely used textbooks (at least in the context of Indonesia). It is easy for student to

understand the concept if the example is continuous force. Previous research (Mansyur *et al*, 2010) showed that not only students, many teachers have difficulty in solving problem that relates impulse force. The common examples of the concept from researches result were the crashing of two objects (a car crashes a truck; an apple hits the Earth; etc). If we ask about the magnitude of the forces, for example: "where is object that 'feels' greater force" or "where is object gives force greater than the other?". Generally, they referred to mass, velocity, size or combination of mass and velocity of the objects. "Faster object or massive object gives greater force to other" is common statement.

Research about the Newton's third law was also conducted by Bao *et al* (2002). They concluded that students generally have inappropriate reason about the magnitude of interaction forces of two objects concerning with velocity, mass, pushing, and acceleration.

A classic example is given in Elby's paper introducing the Elby's pair (Elby, 2001). Students are well known to have difficulty believing in Newton's third law. Although they can often state the law (especially the "action-reaction" form), they often either don't know what the words mean or don't believe that the law applies widely (Redish, 2004).

Situation of an object (mass M) moves in the certain velocity and it collides another object (mass m , $m < M$) can involve abstract primitive reasoning that "greater agent" results "great effect" is well-known as *facet*. Facets may represent consistently applied explanations manifested in a declarative knowledge. They can also express certain strategies, elements of students' characteristic behavior (procedural knowledge), when coping with particular questions and problems. Facets are more context specific, and thus less fundamental than *p-prims*. Facets may incorporate several concepts, related in such a way as to represent individual comprehension of the situation. A facet could be a generic bit of knowledge, specific context of reasoning or could express certain strategies (Galili and Hazan, 2000). The example of a generic bit of knowledge is expression "more means more".

The idea is challenge for teacher or lecturer in teaching concept of the given phenomena. By including context of impulse force in assessing conceptual knowledge related to Newton's third law, educators could consider the context in their teaching activity.

Our research included development of the test and analysis of scientific and representational consistency of first year physics education students

at a university by using the test. In this paper, we presented description of development process, aspects of relevant concept, and isomorphic features of the test.

II. Method

2.1 Procedure of Development of the Test

Process of the test development consisted of six main stages, included: constructing draft, expert judgment, revision, tryout, analyses of test items, and re-revision.

Stage-1: Draft construction

In this stage, we considered results of relevant researches, theoretical aspects of a test, characteristics of isomorphic items, and representational consistency. Based on the considerations, we designed initial draft of the test. The initial draft consisted of 30 items that divided into three main groups: verbal, diagram/vectorial, and graphical representation. In this context, we obtained 10 items for each representation. Each item in verbal group has one item that equivalent to one item in diagram/vectorial and one item in graphical group. There is a similarity of context in their stems, but different in form of option's representation. The contexts of test items include: gravitation, electrostatics, magnetic, pushing, and crash (impulse force). We adopted some items from Nieminen *et al*. (2010), Bao *et al*. (2002), and Force Concept Inventory (FCI) (Jackson, 2009) with some modification. Equivalent items in different representation for the adopted items were constructed by authors. For example, if we adopted one item from FCI or others in verbal representation, we constructed one item for diagram/vectorial and one item for graphical representation or otherwise in the similar context. From the step, we obtained variant of isomorphic items in three representation formats (Figure 1). In this stage, we obtained Draft-1.

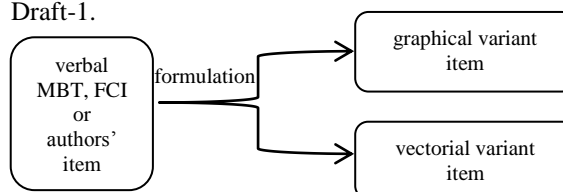


Fig.1. Formulation of test from verbal MBT, FCI or authors item to other representations.

Stage-2: Expert Judgment

Expert judgment was conducted to know the scope and appropriateness of concept, construction and

suitability of the test in relation to principles aspect in constructing a test. For this purposes, we engaged two physics lecturers from a university to give opinion about the aspects. We also asked three undergraduate students and three postgraduate students (they are junior or high school teachers) to give information about clarity of the stem, diagram, graph, and options of test items.

Stage -3: Revision

Based on the activity in the Stage-2, we revised the draft by considering the experts', teachers', and students' suggestions. The revision included: language use, concept aspect, structure of stem, the sequence of options (word, sentence, or number), and clarity of graph or diagram. In this stage, we obtained Draft-2.

Stage -4: Try Out

Try out was conducted on 23 second year physics education students. They have enrolled Basic Physics-I and Basic Physics II courses in first and second semester. Basic Physics-I course includes a general introduction to physics, elementary kinematics and Newton's laws. Basic Physics-II course included electrostatic and electrodynamics (Coulomb's law, etc) and magnetism. In this stage, beside we obtained the chosen options and scores of the students, we also received participants' opinion about the clarity of the test items.

Stage -5: Analyses of test item

For analyzing the test and its items, we used Anates v.4.0 [6]. Input data of the program includes the chosen option by participants and answer key for each item. Output data includes discrimination and difficulty index, item validity and reliability.

Stage -6: Re-revision

Based on the Stage-4 and Stage-5, we conducted revision on Draft-2. Our focuses were test construction and clarity of language. From this stage, we obtained Draft-Final (not attached).

2.2 Data Analysis

Data analysis included the overall aspects of multiple-choice tests and their items. In this study, we referred four measures based on classical test theory. Three of them were for item analyses: item difficulty (P), discrimination index (D), and point biserial coefficient (r_{pbi}). A measure for test analysis: Kuder-Richardson reliability index (r_{test}). The four aspects of test and its items were determined by using Anates v.4.0 (Karnoto and Wibisono, 2004). Description of the four aspects of a test that used in the software is briefly explained in the following.

2.2.1 Item Difficulty Index

The item difficulty index (P) indicates the difficulty of a certain test item. The value of the difficulty index varies between 0 and 1, with 0.5 being the best value. We used range for acceptable values is $0.30 \leq P < 0.70$ (Arikunto, 2002).

2.2.2 Item Discrimination Index

The item discrimination index (D) is a measure of the discriminatory power of an item. It indicates how well an item differentiates between high-achieving and low achieving students. The simplest and most often used system to categorize students into high- and low-achieving groups is to divide them in two equal-sized groups based on the median of the students' total score (Nieminen *et al.*, 2010). We used values of $D > 0.20$ have been considered acceptable (Arikunto, 2002).

2.2.3 Point Biserial Coefficient

The point biserial coefficient indicates how consistently an item measures students' performance in relation to the whole test. The desirable value for the point biserial coefficient is $r_{pbi} \geq 0.4$ for $df = 21$, $p = 0.05$ (Karnoto and Wibisono, 2004).

2.2.4 Kuder-Richardson Reliability Index

KR-20 (r_{test}) is an often used measure of internal consistency when test items are dichotomous (i.e., correct or incorrect). If a test has good internal consistency, different test items measure the same characteristic, and there are high correlations between individual test items. The values of r_{test} range from 0 to 1. A widely used criterion for a reliable group measurement is $r_{test} \geq 0.70$.

2.2.5 Concentration Analysis

As a way to validate the effectiveness of this multiplechoice instrument, we used the *Concentration Analysis* (Bao, 2002) to evaluate the design of the distracters. The way in which the students' responses are distributed on research-based multiple-choice questions can yield information on the students' state. This measure is defined as the concentration factor, C , which is a function of students' responses and takes a value in $[0,1]$. Larger values represent more concentrated responses with 1 being a perfectly correlated response and 0 a random response. This concentration factor can be calculated with Eq. (1):

$$C = \frac{\sqrt{m}}{\sqrt{m}-1} \left(\frac{\sqrt{\sum_{i=1}^m n_i^2}}{N} - \frac{1}{\sqrt{m}} \right) \quad (1)$$

where m represents the number of choices for a particular question, N is the number of students, and n_i is the number of students who select choice i of the question.

2.3 RESULT AND ANALYSIS

In this section, results of research is presented by focusing on the characteristics of the test, including construction, isomorphic aspects, and theme of items. A sample of test item is also presented in this section. As an example, construction of test could be seen in Figure 2 (translated from Indonesian). In the context of magnetic theme, item 8, 18 and 28 have corresponding multiple-choice alternative as description of isomorphic aspect of the test. Each alternative of item 8 has pair in alternative of item 18 and 28 in different format. The representational formats of the alternatives are a bar chart (item 8), vectorial (item 18) and verbal (item 28). The questions of bar chart and vectorial items include explanation of notations in the related items. Distribution and numbering of items test based on relationship of representation formats is presented in Table 1.

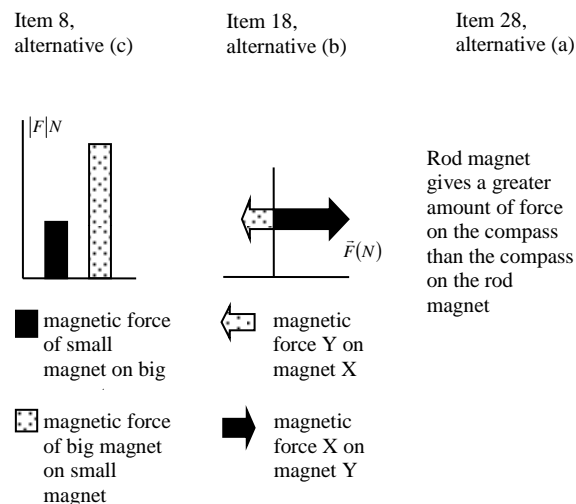


Fig. 2. Corresponding multiple-choice alternatives of magnetic theme in the test

Table 1. Distribution of the themes and corresponding items based on formats

Theme/Context	Representation/Item Number		
	Verbal	Graphical	Diagram
Grav.-1	1	11	21
Electr.-1	22	2	12
Mag.-1	13	23	3
Exertion-1	4	14	24
Crashing-1	25	5	15
Grav.-2	16	26	6
Electr.-2	7	17	27
Mag.-2	28	8	18
Exertion-2	19	29	9
Crashing-2	10	20	30

The data analysis results of try out based on classical test theory is presented in Table 2 and the summary of the analyses is in Table 4.

The difficulty index (P), discrimination index, and point biserial coefficients values for each item of the test are shown in Table 2. The values of P vary between 0.22 and 0.70. Only six items were below 0.30. The averaged difficulty index is 0.42. The discrimination index values of items ranged from 0.00 to 1.00. The desired value of 28 items were above 0.20. Hence, the majority of items of the test had quite satisfactory discriminatory power. The averaged discrimination index was 0.62, which was also in the satisfactory range.

Table 2. Items, themes/contexts and results of item analysis

No	Theme/Context	Repr.	P	D	r_{phi}	C
1	Grav.-1	Verb	0.22	0.83	0.84	0,07
2	Electr.-1	Graph	0.48	0.67	0.51	0,13
3	Mag.-1	Diag	0.22	0.17	0.14	0,13
4	Exert-1	Verb	0.39	0.67	0.67	0,16
5	Crash-1	Graph	0.52	0.83	0.64	0,26
6	Grav.-2	Diag	0.26	0.67	0.72	0,17
7	Electr.-2	Verb	0.26	0.50	0.52	0,33
8	Mag.-2	Graph	0.35	0.33	0.44	0,12
9	Exert-2	Diag	0.61	0.67	0.53	0,23
10	Crash-2	Verb	0.61	0.50	0.48	0,25
11	Grav.-1	Graph	0.70	0.33	0.25	0,30
12	Electr.-1	Diag	0.39	0.67	0.65	0,23
13	Mag.-1	Verb	0.26	1,00	0.84	0,21
14	Exert-1	Graph	0.61	0.67	0.56	0,24
15	Crash-1	Diag	0.35	0.67	0.62	0,07

16	Grav.-2	Verb	0.61	0.67	0.36	0,22
17	Electr.-2	Graph	0.52	0.83	0.62	0,18
18	Mag.-2	Diag	0.43	0.67	0.48	0,19
19	Exert-2	Verb	0.61	0.50	0.45	0,26
20	Crash-2	Graph	0.35	0.67	0.72	0,28
21	Grav.-1	Diag	0.26	0.50	0.37	0,17
22	Electr.-1	Verb	0.30	0.83	0.76	0,07
23	Mag.-1	Graph	0.43	0.50	0.461	0,22
24	Exert-1	Diag	0.39	1.00	0.83	0,09
25	Crash-1	Verb	0.35	1.00	0.84	0,07
26	Grav.-2	Graph	0.61	0.00	0.06	0,23
27	Electr.-2	Diag	0.39	0.83	0.53	0,14
28	Mag.-2	Verb	0.30	0.33	0.47	0,19
29	Exert-2	Graph	0.39	0.83	0.70	0,09
30	Crash-2	Diag	0.48	0.50	0.57	0,26

The averaged value of point biserial coefficient was 0.55. They were above 0.4 except for four items, which supports the notion that almost all the items of the test are reliable and consistent. The point biserial coefficients indicate the items are eligible to measure students' performance in relation to the whole test. For items were below the criteria, we revised them by focusing on scope of concept, context and clarity of items' stem and options.

The averaged value of concentration was 0.19. It was below 0.2, which shows the distribution of respondents' choice and the function of distracters. Reliability index of the test was 0.95. The value shows the test reliable to use in collecting data.

Tabel 3 Distribution of concentration category for each theme

Theme	Number items based on concentration category			C_{av}	Categ.
	High	Mod.	Low		
Gravitation	0	3	3	0.19	Low
Electrostatic	0	2	4	0.18	Low
Magnetic	0	2	4	0.18	Low
Exertion	0	3	3	0.18	Low
Crashing	0	4	2	0.20	Low

Table 4. Summary of evaluation result of test

Evaluation measure	Values of the Test	Desired values
P	Average of 0.42	$0.30 \leq P \leq 0.70$
D	Average of 0.62	> 0.20
r_{pbi}	Average of 0.55	≥ 0.40 ($p=0.05$)
r_{test}	0.95	≥ 0.70
C_{av}	0,19	$< 0,20$

The test has role in probing the scientific and representational consistency. How the test examine the consistencies. We followed (Nieminen *et al.*, 2010) to categorize the level of scientific and representational consistency. In this paper, we just describe the role of the test in examining the consistencies. Students exhibited representational consistency when all the answers in a given theme were consistently correct or consistently incorrect. Furthermore, students exhibited scientific consistency when all the answers in a given theme were correct in terms of both physics and representations. In this analysis, scientific consistency is considered a sub concept or a special case of representational consistency. For both representational and scientific consistency, students' answers in a given theme were graded in the following way:

- Two points, if they had chosen corresponding alternatives in all three items of the theme.
- One point, if they had chosen corresponding alternatives in two of the three items of the theme.
- Zero points, if no corresponding alternatives in the items of the theme were selected

In order to evaluate students' scientific consistency and representational in the whole test, the average points for all the themes were calculated. This meant that a student's points for ten themes were added together and divided by ten, so the average was also between zero and two points. On the basis of the average points, students' scientific and representational consistency was categorized into three levels as presented in Table 5.

Table 5. Categorization of consistency

Level	Value	Category
I	$1.7 < \text{Average}$ (85% of the maximum) or higher	consistent
II	$1.2 \leq \text{Average} \leq 1.7$ (60%–85% of the maximum)	moderately consistent
III	$\text{Average} < 1.2$	inconsistent

The categorization rules are arbitrary, but they are similar to those used with the FCI. An FCI score of 60% is regarded as being the 'entry threshold' to Newtonian physics, and 85% as the "mastery threshold" (Nieminen *et al.*, 2010 and Jackson, 2009).

Beside for investigating the scientific and representational consistency, the test could be used to diagnose the alternative conceptions and the activation of cognitive element such as facet of knowledge. Bao *et al* (2002) stated that successful instruction should also include effective assessment tools to provide accurate and context-rich information of students' state of understanding. The test could be a useful assessment tool in research

and instruction. In the context of score-based methods, it has several advantages: (1) It uses multiple-choice instruments making it appropriate and feasible to implement this method in large classes; (2) The probing instruments and analysis methods are based on systematic research of student conceptual models and thus can provide detailed and validated information on the state of student understanding.

III. Conclusion and Recommendation

Result of the analyses showed that the overall test items have eligibility as a good test to measure various aspects of students' understanding of concept in relation to the scientific and representational consistency. The averaged values of the test were in the satisfactory range. There is a limitation of the test related to scope of the concept. The overall statements of correct answers of test items were similar. For instant, if a respondent refers to Newton's third law or general statement "force by A on B is equal to B on A" as main answer key for all items, then it is possible for the respondent to answer the test with all correct choices. For this condition, it is difficult to interpret his/her representational consistency. We may interpret it that he/she is consistent in scientific, but there is no guarantee that he/she is consistent in representational aspect. It is needed further study to develop tests that could explore students' consistency related to other concepts / contexts.

Acknowledgement

We would like to thank to General Directorate of Higher Education, Ministry of National Education and Culture of Indonesia for funding this research in Hibah Tim Pascasarjana scheme under contract No. 172/UN28.2/PL/2013. Special thanks for anonymous experts have validated the test.

IV. References

- Arikunto, S. 2002. *Dasar-Dasar Evaluasi Pendidikan* (Edisi Revisi). PT. Bumi Aksara. Jakarta.
- Bao, L. Hogg, K. and Zollman, D. 2002. Model Analysis of Fine Structures of Student Models: An Example with Newton's Third Law. *American Journal Physics*. Vol. 70, (7).
- Cock, M.D. 2012. Representation Use and Strategy Choice in Physics Problem Solving. *Physical Review Special Topics-Physics Education Research*, 8, 020117.
- Elby, A. 2001. Helping physics students learn how to learn. *Phys. Educ. Res. Suppl. to Am. J. Phys.* 69 S54-S64.
- Galili, I. and Hazan, A. 2000. The Influence of an Historically Oriented Course on Students' Content Knowledge in Optics Evaluated by Means of Facets-Schemes Analysis. *American Journal of Physics*, 68 S1, S3-S15.
- Halliday, D. and Resnick, R. 1994. *Fisika* (Penerjemah: Silaban, P dan Sucipto, E). Erlangga. Jakarta.
- Jackson, J (jane.jackson@asu.edu). 2009. Revised-Force Concept Inventory (D. Hestenes, M. Wells, and G., Swackhamer, 1995). (2009, 03 April), e-mail to Jusman Mansyur (jushmanyurfis@yahoo.com).
- Karnoto and Wibisono. Y. 2004. Anates versi 4.0.
- Kohl, P. B. and Finkelstein, N. D. 2006. Effects of Representation on Students Solving Physics Problems: A Fine-Grained Characterization. *Physical Review Special Topics-Physics Education Research*, 2, 010106.
- Mansyur, J., Setiawan, A. & Liliarsari. 2010c. Model Mental Siswa, Mahasiswa dan Guru pada Hukum III Newton dalam Konteks Problem Solving: Kasus Gaya Impuls. *Prosiding Seminar Nasional Pendidikan*, Bandar Lampung, 27 Februari 2010. Bandar Lampung: Universitas Lampung.
- Niemenen, P., Savinainen, A. and Viiri, J. 2010. Force Concept Inventory-Based Multiple-Choice Test for Investigation Student's Representational Consistency. *Physical Review Special Topics-Physics Education Research*, 6, 020109.
- Redish, E.F. 2004. A Theoretical Framework for Physics Education Research: Modeling Student Thinking, in E. Redish and M. Vicentini (Eds.), *Proceedings of the Enrico Fermi Summer School, Course CLVI* (Italian Physical Society, 2004).
- Tipler, P.A. 1998. *Fisika* (Penerjemah: Prasetyo, L dan Adi, R.W). Erlangga. Jakarta.