



The Three-tier Diagnostic Instrument: Using Rasch Analysis to Develop and Assess K-10 Students' Alternative Conceptions on Force Concept

Nuzulira J. Fratiwi*, Taufik R. Ramalis and Achmad Samsudin

Department of Physics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

*Corresponding author, E-mail: nuzulira.janeusse.fratiwi@student.upi.edu

Abstract

The aim of this study was to assess K-10 students' alternative conceptions on force concept using Three tier-Force Concept Inventory (T-FCI) and develop a four-tier diagnostic instrument from three-tier using Rasch analysis. The development of the four-tier test from the three-tier test was done because they still cannot completely distinguish the confidence selections for the key response from confidence selections for reasoning. The research method used the 4D model (Defining, Designing, Developing and Disseminating). The sample of this study involved 36 K-10 students (12 boys and 24 girls, their average age about 16 years old). At the developing stage, 13 items of the Force Concept Inventory were developed in the form of a three-tier test utilizing the Three tier-Force Concept Inventory (T-FCI). This is aimed to get the students' reasoning for the second development of FCI in the form of a four-tier test, which is named Four tier-Force Concept Inventory (F-FCI). The developing of a four-tier test is based on students' answer on the T-FCI and analyzing with the Rasch model. Based on analysis of the Rasch model, the T-FCI were valid and reliable to use and one item was unusable. At the disseminating stage, students' conceptions were categorized as five varieties as Sound Understanding (SU) (1%), Partial Understanding (PU) (27%), Alternative Conceptions (AC) (60%), No Understanding (NU) (2%) and No Coding (NC) (10%). This study can be a reference for researchers to further research and teachers to prepare a learning activity that can reduce students' alternative conceptions.

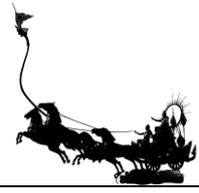
Keywords: *Three-tier diagnostic instrument, Four-tier diagnostic instrument, Students' alternative conceptions, Force concepts*

1. Introduction

Many researchers around the world have been disputing that the most imperative trait that students convey to their lessons are conceptions, the greatest of which has different views from various experts (Samsudin et al., 2017; Carmona, Criado, & Guzman, 2018). An unscientifically adequate concept assumed through students known as 'alternative conceptions' (Fariyani, Rusilowati, & Sugianto, 2017; Fratiwi et al., 2017; Liu & Fang, 2016; Kose, 2008). Kaltakci and Didis (2007) recorded the belongings of alternative conceptions as: (1) powerfully apprehended, unchanging cognitive assemblies (2) differ from scientists' conceptions (3) disturb how students comprehend natural occurrences and scientific descriptions (4) must be removed. Thus, students' alternative conceptions can be robust and impervious to modify with scientific ones (Fratiwi, Samsudin, & Costu, 2018; Hermita et al., 2017; Oliver et al., 2017; Larkin & Jorgensen, 2016).

Alternative conceptions are an obstacle for students when they are attempting to recognize the concepts of physics (Pujayanto et al., 2018). For the reason that alternative conceptions can struggle in the process of integrating novel information, the alternative conceptions must be measured as presently as possible (Hermita et al., 2017; Canu, Hosson, & Quque, 2016). Accurate diagnosing of students' alternative conceptions has come to be the main primary phase in the direction of improvement an accepting of student knowledge (Kaltakci-Gurel, Eryilmaz, & McDermott, 2017). Students' alternative conceptions can be detected through a diagnostic instrument (Fazio & Battaglia, 2018). The different diagnostic instruments have been advanced and as well as practices by researchers. Nevertheless, a separately diagnostic instrument has some benefits as well as weaknesses (Kaltakci-Gurel et al., 2017; Kaltakci-Gurel, Eryilmaz, & McDermott 2015).

One of the diagnostic instruments is a four-tier test that has been the adaptation of the three-tier form into a four-tier format (Caleon & Subramaniam, 2010). The modification of the three-tier test still cannot completely distinguish the confidence selections for the key response (first tier) from confidence



selections to reasoning (second tier). Therefore, it may miscalculate students' scores and undervalue their lack of knowledge (Kaltakci-Gurel et al., 2015). In the four-tier test, the first tier of it is a conventional multiple-choice test which its distractors are talking about in detailed alternative conceptions. The second tier of the test requires the confidence of the response in the first tier. The third tier of the test requires for the reasoning for the response in the first tier. And then, the fourth tier of the test requires for the confidence of the response in the third tier. The four-tier test is more accurate to diagnose students' alternative conceptions (Fратиwi et al., 2017; Gurel et al., 2017; Lin, 2016). Four-tier tests can correctly evaluate students' alternative conceptions which are allowed in no understanding and partial understanding.

In physics education, there are many standard tests to assess students' conceptions such as the Force Concept Inventory (FCI), Rotational Kinematics Inventory (RKI), Electromagnetics Concept Inventory (EMCI) and others. Force Concept Inventory (FCI) has been developed by Hestenes, Wells, and Swackhamer (1992) about force concept. Force concept is the fundamental concepts in physics, and accordingly it important for students to absorb further physics' concepts (Fратиwi et al., 2018). The existing FCI has a multiple choice form that can be directed in enormous samples. However, it cannot be afforded in a complex investigation about students' replies and may not be owing to the alternative conceptions, but might be a mistaken response with truthful reasoning (Kaltakci-Gurel et al., 2017; Kaltakci & Didis, 2007). Therefore, we use the three-tier diagnostic instrument (was developed from FCI) to assess K-10 students' alternative conceptions as an initial study. Moreover, the four-tier diagnostic instrument was developed from the three-tier diagnostic instrument using Rasch analysis.

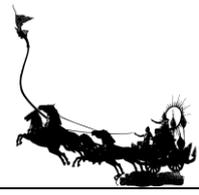
2. Objectives

The objectives of this study were the following:

1. To assess students' alternative conceptions about force concept using Three tier-Force Concept Inventory (T-FCI)
2. To develop the diagnostic instrument in the form of a four-tier test from Three tier-Force Concept Inventory (T-FCI)

3. Materials and Methods

The research method utilized 4D model (Defining, Designing, Developing, and Disseminating). The 4D model is an unpretentious tuition design model that helps the researchers to design a product (Irawan, Padmadewi, & Artini, 2018). At the defining and designing stage, the three-tier and four-tier test were defined and designed. The three-tier diagnostic instrument is a diagnostic test with three tiers. The first tier is the original multiple choice, the second tier has reasoning choices and the third tier is confidence rating of the first and second tier as "Sure" and "Not Sure". The four-tier diagnostic instrument is a diagnostic test with four tiers. The first tier of a four-tier test is multiple choice, the second tier is the confidence rating of the answer on the first tier, the third tier has reasoning choices of the answer on the first tier, and then the fourth tier is the confidence rating of the reasoning choice on the third tier. From this definition, the design of the three-tier and four-tier test is shown in Figure 1 and Figure 2.



1.1	Question Answer choice: a. b. c. d. e.
1.2	Your reasons regarding the answer choices in question 1.1: (reason)
1.3	Are you sure of the reason answer to question 1.1 and 1.2? (confidence rating) a. Sure b. Not sure

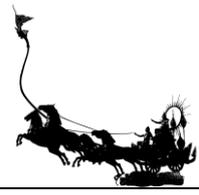
Figure 1 The design of the three-tier diagnostic instrument

1.1	Question Answer choice: a. b. c. d. e.
1.2	Are you sure of the answer to question 1.1? (confidence rating) a. Sure b. Not sure
1.3	Your reasons regarding the answer choices in question 1.1: (reason) a. b. c. d. e. (empty)
1.4	Are you sure of the reason answer to question 1.3? (confidence rating) a. Sure b. Not sure

Figure 2 The design of the four-tier diagnostic instrument

The number of 1.1 signify to an item or question number one for the first tier, the number of 1.2 indicates to item number one for the second tier, and so on until 13 items.

At the developing stage, the three-tier test was developed from 13 original items of the Force Concept Inventory (FCI) named Three tier-Force Concept Inventory (T-FCI). The 13 items from 30 items of FCI adjusted to subject matter in a senior high school at Indonesia. This first development aimed to explore students' reasoning using open answer at the second tier on a three-tier test. The example of T-FCI shown is in Figure 3.



2.1 A large truck collides head-on with a small compact car. During the collision:

- a. the truck exerts a greater amount of force on the car exerts on the truck.
- b. the car exerts a greater amount of force on the truck that the truck exerts on the car.
- c. neither exerts a force on the other, the car gets smashed simply because it gets in the way of the truck.
- d. the truck exerts a force on the car but the car does not exert a force on the truck.
- e. the truck exerts the same amount of force on the car as the car exerts on the truck.

2.2 Your reasons regarding the answer choices in question 2.1:

.....

.....

.....

2.3 Are you sure of the answer to questions 2.1 and 2.2?

- a. Sure
- b. Not sure

Figure 3 The example of three tier-force concept inventory (T-FCI)

Students' reasoning on T-FCI was organized as the third tier of the four-tier test. The second development process of FCI in the form of a four-tier test was named Four tier-Force Concept Inventory (F-FCI). Moreover, the development of a four-tier test was based on analyzing Rasch model (validity and reliability). The use of the Rasch model can measure precisely because it does not depend on the sample of a particular item or the person selected in an exam (Sumintono & Widhiarso, 2014). The developing process is shown in Figure 4. At the disseminating stage, the result from the primary implementation was an analysis of students' conceptions.

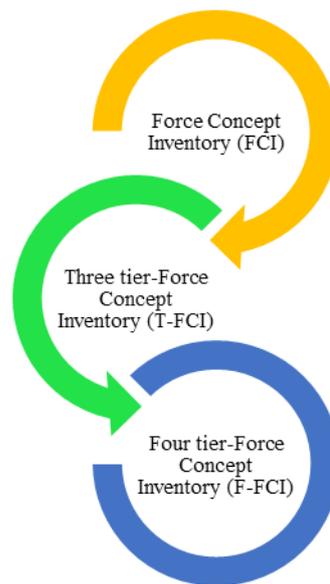


Figure 4 The developing process of four tier-force concept inventory (F-FCI)

The samples in this study were 36 K-10 students (12 boys and 24 girls, their average age about 16 years old). Students took the T-FCI for about 60 minutes. The students' conceptions were categorized into five varieties as the Sound Understanding (SU), Partial Understanding (PU), Alternative Conceptions (AC), No Understanding (NU) and No Coding (NC). These categories were based on students' responses on the T-FCI. The varieties of students' conceptions were modified from the Kaltakci and Didis (2007) category. Students were categorized as Sound Understanding (SU) if they answer correctly (scientific concept) at the



first and second tier, and sure with their answer. For other categories, it can be seen in Table 1. And then, for the calculation needs, students' conceptions were scored as shown in Table 2. If students were categorized as Sound Understanding (SU), and they are given a score of 2 and so on. Students' scores are needed for calculation using Rasch analysis.

Table 1 The five categories of students' conceptions

Category	Tier-1	Tier-2	Tier-3
Sound Understanding (SU)	Correct	Correct	Sure
Partial Understanding (PU)	Correct	Correct	Not sure
	Incorrect	Correct	Not sure
	Correct	Incorrect	Not sure
Alternative Conceptions (AC)	Incorrect	Incorrect	Not Sure
	Correct	Incorrect	Sure
No Understanding (NU)	Incorrect	Correct	Sure
No Coding (NC)	Students do not answer totally or piece of tiers in test items.		

Table 2 The score of students' conceptions

Category	Score
Sound Understanding (SU)	2
Partial Understanding (PU)	1
Alternative Conceptions (AC)	0
No Understanding (NU)	0
No Coding (NC)	0

4. Results and Discussion

For results and discussion, we separated into two parts as assessing (disseminating) and developing the process, as follows:

4.1 Assess K-10 Students' Alternative Conceptions (Disseminating)

Based on the primary study using T-FCI, we have categorized students' conceptions as shown in Figure 5.

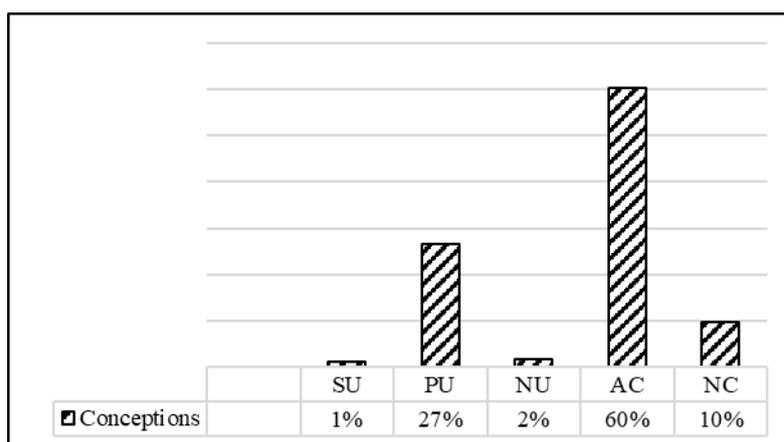


Figure 5 The percentage of each student's conceptions

Grounded in Figure 5, it can be seen the percentage of each student's conceptions. The percentage in the Sound Understanding (SU) category is 1%, Partial Understanding (PU) is 27%, Alternative Conceptions (AC) is 60%, No Understanding (NU) is 2%, and No Coding (NC) is 10%. The highest percentage is in the AC category, while the lowest percentage is in the SU category. This indicates that students' alternative conceptions in the force concept are fairly high. This is in accordance with the research



of Poutot & Blandin (2015) which states that many students hold alternative conceptions after being identified with FCI. Moreover, the percentage of alternative conceptions is higher than other conceptions. This is indicating that the three-tier test cannot accurately diagnose students' conceptions as argued by Kaltakci-Gurel et al (2017). The percentage of students' alternative conceptions for each item was presented in Figure 6.

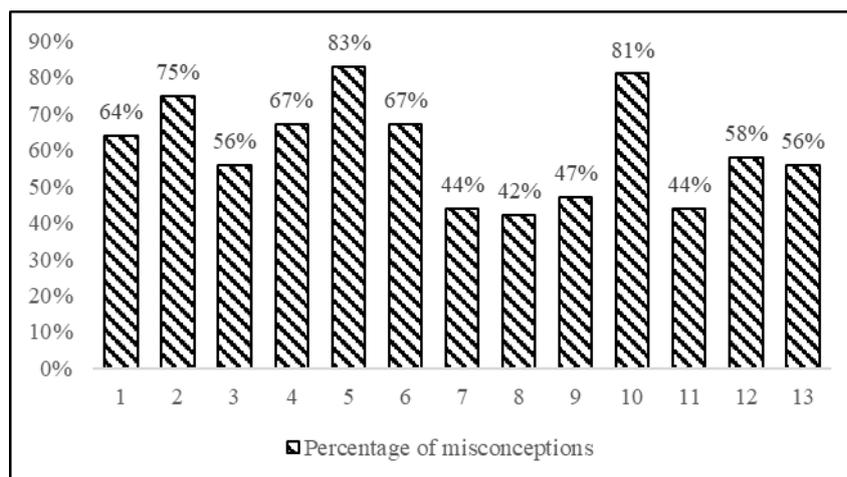


Figure 6 The percentage of students' alternative conceptions for 13 items of T-FCI

In Figure 6, it can be seen that the highest alternative conception is in item number 5 (83%), while the lowest alternative conception is in question number 8 (42%). One example of students' alternative conception in question number 5 is that when a box is given a force on a rough floor, suddenly the given force is removed. Most students answer that the box will stop immediately. This happens because when no force is given, the object that first moves will immediately stop. Students do not associate these occasions with the concepts of acceleration and velocity. When the force is removed, there is only a constant friction force which causes constant deceleration of the object and the velocity will decrease constantly until the object stops. This is in accordance with the findings of Liu & Fang (2016) that students assume when there is no force given then the velocity of the object will be constant (zero).

4.2 Developing Four tier-Force Concept Inventory (F-FCI)

At the developing stage, there is a two development process. Firstly, we developed FCI in the form of three-tier in the Three tier-Force Concept Inventory (T-FCI). The first tier is adopted from FCI, the second tier is an open-ended answer for students' reasoning for the answer at the first tier and the third tier is a confidence rating for students' answer at the first and second tier (as shown in Figure 3). The open-ended answer at the second tier is intended for collective students' answer and we choose three until four answers for the third tier in the four-tier test. Therefore, the reasoning choices of the four-tier test were developed from students' thinking in general. Subsequently with the development of T-FCI, we distributed T-FCI to 36 K-10 students.

Students' answers on T-FCI were identified and students' conceptions are categorized as Sound Understanding (SU), Partial Understanding (PU), Alternative Conceptions (AC), No Understanding (NU) and No Coding (NC). After that, students' conceptions were scored based on Table 2. This score can be calculated using the Rasch model for the validity and reliability of the T-FCI. The T-FCI had a 'fulfilled' validity where the index of raw variance explained by measures was 29.8% (Sumintono & Widhiarso, 2014). The validity of T-FCI was not very good, but still in the fulfilled criteria. This indicates that the instrument can measure what should be measured. For more analysis of the validity of the instrument, misfit order can assist to eliminate items which do not fit. Items did not fit if the value of PT-Measure Corr of an



item is not between the value 0.4 until 0.85 (Sumintono & Widhiarso, 2014). Based on the analysis (as shown in Table 3), item number 1 does not fit because the value of PT-Measure Corr is 0.37. Therefore, item number 1 must be eliminated.

Table 3 The result of misfit order using the Rasch analysis

Entry Number	Total Score	PT-Measure	
		Corr.	Exp.
1	14	.37	.58
11	15	.45	.60
7	16	.54	.61
13	11	.60	.55
9	9	.51	.51
5	4	.44	.39
12	10	.57	.53
6	10	.58	.53
2	7	.54	.48
3	12	.61	.56
4	10	.65	.53
10	4	.54	.39
8	18	.71	.63
Mean	10.8		
S.D.	4.1		

If we describe it into a diagram, it can be shown in Figure 7.

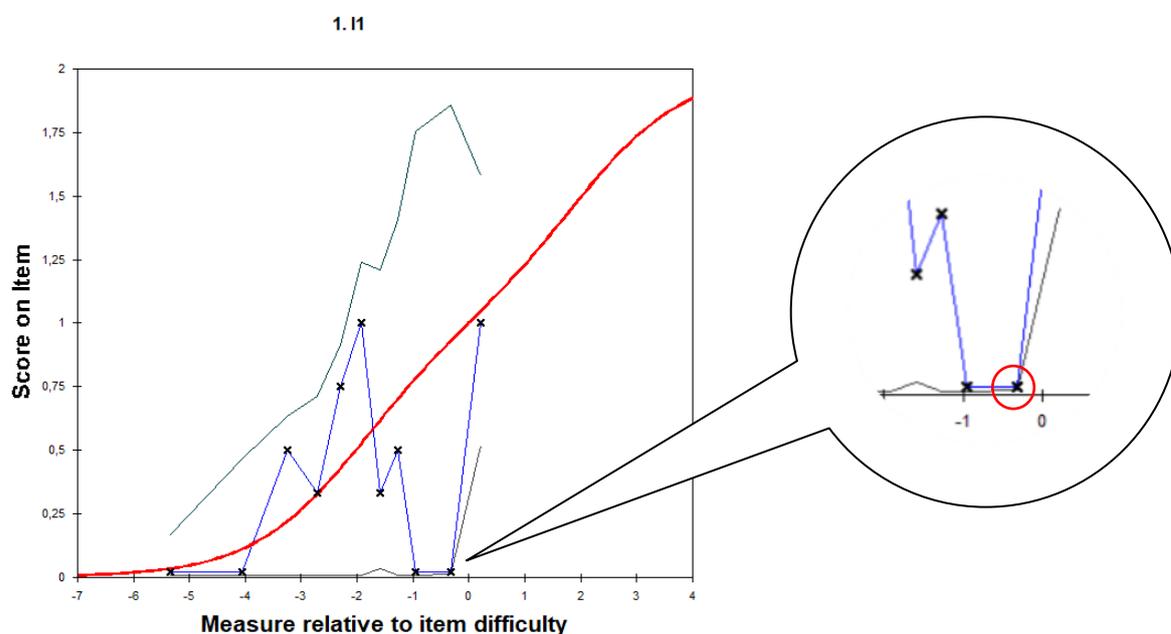


Figure 7 The result of misfit order in form of a diagram

In Figure 7, we can see that one item of T-FCI was on the black line or maybe outside the black line. The cross sign that was outside or at the black line must be eliminated as an analysis from Table 3. The



value of Cronbach Alpha coefficient is 0.82, and designated the relations between 36 students and 13 items. The reliability score is classified as 'very good'. This score recommends that there was a high level of relations between the persons and items (Adams et al., 2018).

From the analysis of the Rasch model, it can be concluded that T-FCI is valid and reliable to be used. The second step was the development of Four tier-Force Concept Inventory (F-FCI) and is based on students' answer at the second tier on T-FCI. After that, based on analysis through the Rasch model, one item (number 1) must be eliminated. Based on the construction, item number 1 does not describe the condition of a phenomenon and the question posed is less clear. This condition makes students not answer if they do not look at the answer choices. However, we just developed 12 items of F-FCI. The example of F-FCI is shown in Figure 8.

- 2.1 A large truck collides head-on with a small compact car. During the collision:

 - a. the truck exerts a greater amount of force on the car exerts on the truck.
 - b. the car exerts a greater amount of force on the truck that the truck exerts on the car.
 - c. neither exerts a force on the other, the car gets smashed simply because it gets in the way of the truck.
 - d. the truck exerts a force on the car but the car does not exert a force on the truck.
 - e. the truck exerts the same amount of force on the car as the car exerts on the truck.

2.2 Are you sure of the answer to questions 2.1?

 - a. Sure
 - b. Not sure

2.3 Your reasons regarding the answer choices in question 2.1:

 - a. The force given is equal but in the opposite direction and not influenced by the mass of the object.
 - b. The force given is influenced by the mass of the object, the greater mass then the greater force applied.
 - c. The force given is influenced by the volume of the object, the greater volume then the greater force applied.
 - d. The force given is equal when the objects are moving.
 - e.

2.4 Are you sure of the answer to questions 2.3?

 - a. Sure
 - b. Not sure

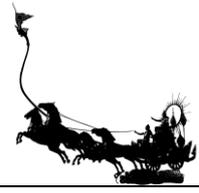
Figure 8 The example of four tier-force concept inventory (F-FCI)

5. Conclusion

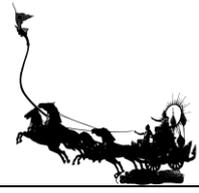
Students' alternative conceptions can be detected through Three tier-Force Concept Inventory (T-FCI). Based on T-FCI, the percentage in the Sound Understanding (SU) category is 1%, Partial Understanding (PU) is 27%, Alternative Conceptions (AC) is 60%, No Understanding (NU) is 2%, and No Coding (NC) is 10%. The highest percentage is in the AC category, while the lowest percentage is in the SU category. This indicates that students' alternative conceptions in the force concept are fairly high. The four-tier diagnostic instrument was developed from the three-tier test. At the developing stage, there is a two development process for T-FCI and F-FCI, which is based on the defining and designing stages. From Rasch analysis, the T-FCI was valid and reliable to use, and then one item of T-FCI must be eliminated. Based on the research findings, a four-tier diagnostic instrument can be developed. This study can be a reference for researchers to further research and teachers to prepare a learning activity that can reduce students' alternative conceptions.

6. References

- Adams, D., Sumintono, B., Mohamed, A. & Noor, N. S. M. (2018). E-learning readiness among students of diverse backgrounds in a leading Malaysian higher education institution. *Malaysian Journal of Learning and Instruction*, 15(20), 227-256.
- Caleon, I. S. & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *International Journal of Science Education*, 32(7), 939-961.



- Canu, M., Hosson, C., & Duque, M. (2016). Students' understanding of equilibrium and stability: The case of dynamic systems. *International Journal of Science and Mathematics Education*, 14(1), 101-123.
- Carmona, A., Criado, A., & Guzman, M. (2018). Prospective primary teachers' prior experiences, conceptions, and pedagogical valuations of experimental activities in science education. *International Journal of Science and Mathematics Education*, 16(2), 237-253.
- Fariyani, Q., Rusilowati, A., & Sugianto. (2017). Four-tier diagnostic test to identify misconceptions in geometrical optics. *Unnes Science Education Journal*, 6(3), 1724-1729.
- Fazio, C. & Battaglia, O. R. (2018). Conceptual understanding of Newtonian mechanics through cluster analysis of FCI student answers. *International Journal of Science and Mathematics Education*, 17, 1-21.
- Fратиwi, N. J., Kaniawati, I., Suhendi, E., Suyana, I., & Samsudin, A. (2017). The transformation of two-tier test into four tier test on newton's laws concepts. *AIP Conference Proceedings*, 1848, 1-5.
- Fратиwi, N. J., Samsudin, A. & Costu, B. (2018). Enhancing K-10 students' conceptions through computer simulations-aided PDEODE*E (CS-PDEODE*E) on Newton's laws. *Jurnal Pendidikan IPA Indonesia*, 7(2), 214-223.
- Hermita, N., Suhandi, A., Syaodih, E., Samsudin, A., Isjoni, Johan, H., Rosa, F., Setyaningsih, R., Sapriadi, & Safitri, D. (2017). Constructing and implementing a four tier test about static electricity to diagnose pre-service elementary school teacher' misconceptions. *Journal of Physics: Conference Series*, 895, 1-6.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30, 141-158.
- Irawan, A. G., Padmadewi, N., & Artini, L. P. (2018). Instructional materials development through 4D model. *SHS Web of Conferences*, 42, 1-4.
- Kaltakci, D. & Didis, N. (2007). Identification of pre-service physics teachers' misconceptions on gravity concept: A study with a 3-tier misconception test. *AIP Conference Proceedings*, 899, 499-500
- Kaltakci-Gurel, D., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science & Technology Education*, 11, 989-1008.
- Kaltakci-Gurel, D., Eryilmaz, A., & McDermott, L. C. (2017). Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics. *Research in Science & Technological Education*, 1-23.
- Kose, S. (2008). Diagnosing student misconceptions: Using drawings as a research method. *World Applied Sciences Journal*, 3(2), 283-293.
- Larkin, K. & Jorgensen, R. (2016). 'I hate maths: why do we need to do maths?' using ipad video diaries to investigate attitudes and emotions towards mathematics in year 3 and year 6 students. *International Journal of Science and Mathematics Education*, 14(5), 925-944.
- Lin, J. (2016). Do skilled elementary teachers hold scientific conceptions and can they accurately predict the type and source of students' preconceptions of electric circuits?. *International Journal of Science and Mathematics Education*, 14(2), 287-307.
- Liu, G. & Fang, N. (2016). Student misconceptions about force and acceleration in physics and engineering mechanics education. *International Journal of Engineering Education*, 32(1), 19-29.
- Oliver, J. S., Hodges, G. W., Moore, J. N., Cohen, A., Jang, Y., Brown, S. A., Kwon, K. A., Jeong, S., Raven, S. P., Jurkiewicz, M., & Robertson, T. P. (2017). Supporting high school student accomplishment of biology content using interactive computer-based curricular case studies. *Research in Science Education*, 1-26.
- Poutot, G. & Blandin, B. (2015). Exploration of students' misconceptions in mechanics using the FCI. *American Journal of Educational Research*, 3(2), 116-120.
- Pujayanto, P., Budiharti, R., Adhitama, E., Nuraini, N. R. A., & Putru, H. V. (2018). The development of a web-based assessment system to identify students' misconception automatically on linear kinematics with a four-tier instrument test. *Physics Education*, 53, 1-8.



- Samsudin, A., Fratiwi, N. J., Kaniawati, I., Suhendi, E., Hermita, N., Suhandi, A., Wibowo, F., Costu, B., Akbardin, J., & Supriyatman, S. (2017). Alleviating students' misconceptions about newton's first law through comparing PDEODE*E tasks and POE tasks: Which is more effective?. *The Turkish Online Journal of Educational Technology, Special Issue for INTE 2017*, 215-221.
- Sumintono, B. & Widhiarso, W. (2014). Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [*Application of rasch modelling in social science research*]. Cimahi: Trimkom Publishing House