International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

Using Rasch Analysis to Develop Multi-representation of Tier Instrument on Newton's law (MOTION)

*[1]Achmad Samsudin, [2]Nuzulira Janeusse Fratiwi, [3]Taufik Ramlan Ramalis, [4]Adam Hadiana Aminudin, [5]Bayram Costu, [6]Muhammad Nurtanto

Abstract--- *The purpose of this research was to develop Multi-representation of Tier Instrument on Newton's law (MOTION) using Rasch analysis. This instrument can be used to diagnose students' conceptions, especially on Newton's laws. Multi-representation of tier instrument also can be used for three forms of tests (pre-, post- and delayed test), thus students do not realize they have done the same problem. The research method was used ADDIE (Analyse, Design, Develop, Implement and Evaluation) model. The participants were students at one of senior high school at Sukabumi. West Java, as much as 92 students (41 male students and 51 female students, with an average age of 15-17 years) who were in K-11. At evaluating stage, all students' answers were analysed based on the student's conception category, scored, then evaluated using Rasch analysis with Winstep 4.4.5 software. The data analysis including validity and reliability of MOTION. Based on the Rasch analysis, it can be concluded that MOTION was valid and reliable to use, although there are a few questions that need to be minor revised. Other researchers can use MOTION to diagnose students' conceptions of Newton's law material and can also use Rasch analysis to develop a research instrument.*

Keywords--- *Multi-representation, Newton's Law, Rasch Analysis, Tier Instrument, Students' Conceptions*

I. INTRODUCTION

Rasch analysis, distributed in 1960 by Georg Rasch, is a statistical method toward the quantity of human concert, attitudes and insights (Aminudin et al., 2019; Fratiwi et al., 2020; Rasch, 1960; Tesio, 2003). Rasch analysis, grounded on Item Response Theory (IRT) delivers an actual useful different intended for discovering the psychometric belongings of measures for practice in the social sciences (Lamb, Annetta, Meldrum, & Vallett, 2012; Murshed, Phang, Bunyamin, & Binti, 2020; Tesio, 2003; Van Zile-Tamsen, 2017). Rasch analysis is a contemporary alternative technique of extent that generates a measurement stage that matches the standards of an international system of unit where it pieces as a tool through a particular unit of quantity and can attend as an excellent classical (Arsad et al., 2013). Rasch analysis is flattering the most suitable technique for initial investigation in the ground of human sciences where a tool used and

[1]*Department of Physics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia.*
[2]*Department of Physics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia.*
[3]*Department of Physics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia.*
[4]*Department of Physics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia.*
[5]*2Department of Science Education, Yildiz Technical University, Istanbul, Turkey*
[6]*Department of Mechanical Engineering Education, Universitas Sultan Ageng Tirtayasa, Banten, Indonesia*

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

the quantity produces ordinal data (Lamb et al., 2012; Liou & Hung, 2015). Since the Rasch analysis grounded on probabilities, it permits the replies of persons to be projected correctly on entirely substances that suitable the measurement typical, expending only an individual constraint and an item constraint on the similar scale (Perry, 2019; Setiawan, Panduwangi, & Sumintono, 2018; You, Marshall, & Delgado, 2018). Rasch analysis used toward recognizing the most apposite amount of rating scale selections for a student assessment tool of teaching (Boone, Abell, Volkmann, Arbaugh, & Lannin, 2011; Lamb et al., 2012; Van Zile-Tamsen, 2017). In physics education, Rasch analysis has ever approved out for numerous studies (Aminudin et al., 2019; Krell, Redman, Mathesius, Krüger, & van Driel, 2018; Romine & Sadler, 2016). However, studies for multitier tools is still infrequently originated, particularly for multi-representation of tier instrument.

Multi-representation is the explanation of classification or technique through two or more forms (Ainsworth, 2006; Fratiwi, Utari, & Samsudin, 2019; Kurnaz & Arslan, 2014). Concepts represented through verbal sentence script, signs as mathematical forms, images and graphics subsequently that the distribution of information data can take (Hutajulu, Minarti, & Senjayawati, 2019; Laszlo, 2013). Multi-representation can be used to distinguish the students' conceptual understanding. Students who have unstated a concept will not have trouble to direct their understanding in the form of numerous representation (Sutopo & Waldrip, 2014; Theasy, Wiyanto, & Sujarwata, 2018). The capability to use representations is measured as existence a significant for learning physics (Brass, Gunstone, & Fensham, 2003; Kohl, Rosengrant, & Finkelstein, 2007; Sutopo & Waldrip, 2014). Students through advanced representational aptitude have an excellent chance of solving compound problems effectively (Henke & Höttecke, 2015; Malone, 2008; Panaoura, Michael-Chrysanthou, Gagatsis, Elia, & Philippou, 2017; Sutopo & Waldrip, 2014). Several studies have examined the usefulness of multi-representation in teaching (Adadan, 2013). Nonetheless, still infrequently practice multi-representation on a diagnostic test, specifically on tier instrument.

Tier instrument was first advanced by Treagust to recognize students' conceptions as a two-tier test (Chen, Lin, & Lin, 2003; Liampa, Malandrakis, Papadopoulou, & Pnevmatikos, 2019; Sia, Treagust, & Chandrasegaran, 2012; Treagust, 1988). Two-tier tests were defined as diagnostic instruments through the first tier, comprising multiple-choice gratified queries, and the second tier, containing multiple-choice customary of reasons for the answer toward the first-tier (Adadan & Savasci, 2012; Gurel, Eryilmaz, & McDermott, 2015; Treagust, 1986). The limits stated for the two-tier tests envisioned to be recompense by including the third tier to separate items of the test inquiring for the confidence in the responses specific in the first and two tiers (Caleon & Subramaniam, 2010; Peşman & Eryilmaz, 2010). In three-tier tests. The first tier of which involved an ordinary multiple-choice test, the second tier of which was a multiple-choice test question questioning for the reasoning, and the third tier of which was a scale asking for the students' confidence level for the assumed answers for the overhead two. Three-tier tests still have some limits due to the concealed rating of the confidence for the first and second tiers in those tests (Gurel et al., 2015; Kaltakci-Gurel, Eryilmaz, & McDermott, 2017). Meanwhile, there is no linking concerning confidence ratings that demanded the content and reasoning tiers, so the four-tier tests are obtainable more currently. In the four-tier test, the first tier is an unadventurous multiple-choice test, the second tier asks the confidence rating of the response in the first tier, the third tier necessitates the reasoning for the answer in the first tier and the fourth-tier involves the confidence rating of the

International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

answer in the third tier (Fratiwi, Kaniawati, Suhendi, Suyana, & Samsudin, 2017; Kaltakci-Gurel et al., 2017; Kaniawati et al., 2019; Saputra, Setiawan, Rusdiana, & Muslim, 2020).

Presently, the four-tier test is the most actual test for identifying students' conceptions (Adimayuda, Aminudin, Kaniawati, Suhendi, & Samsudin, 2020; Liampa et al., 2019, Samsudin et al., 2019; Samsudin et al., 2020). Kaniawati et al. (2019) were developed diagnostic tests in the form of a four-tier test, namely the Four-Tier Newtonian Test (FTNT). Nevertheless, this test only uses one representation. While we use this test in research to find out the effects of a learning process, students at least during the post-test still remember the questions when pre-test. Therefore, we developed a four-tier test in the form of multi-representation, such as verbal, figure, and mathematical, with the same material (Newton's law), as can be seen in Figure 1. Multi-representation of tier instrument is intending so that students do not comprehend the same problem when working on the pre-test, post-test, or delayed test for research that concerns the learning process.

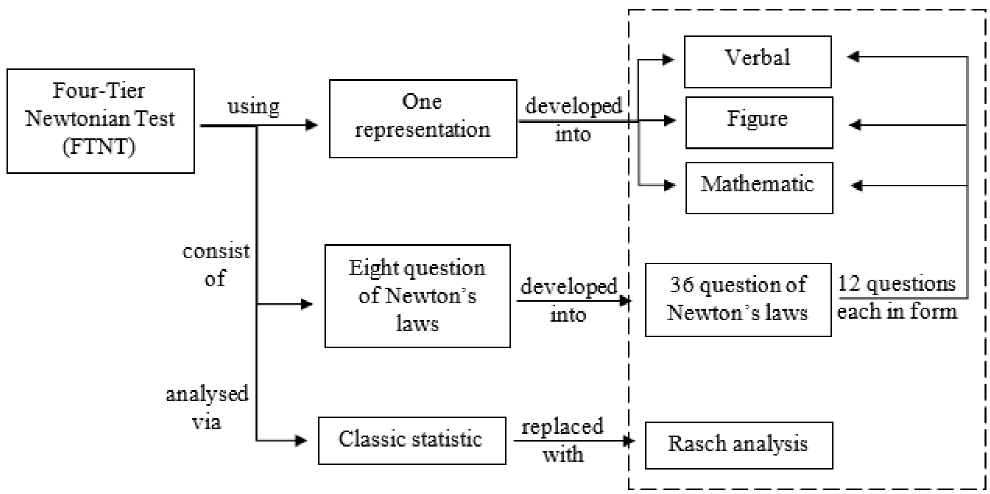


Figure 1: Multi-representation of Tier Instrument Development Process

Multi-representation of tier instrument comprises concepts in Newton's law material such as inertia, the effect of force on velocity, gravitational force, frictional force, action-reaction force and others. The choice of Newton's law is because its concepts are the basic concepts for studying physics (Hermanto, Muslim, Samsudin, & Maknun, 2019; Sutopo & Waldrip, 2014; Velentzas & Halkia, 2013; Yuruk, Beeth, & Andersen, 2009). Newton's laws are imperious since they have readily noticeable plans in the ordinary survives of students (Özcan & Bezen, 2016; Saglam-Arslan & Devecioglu, 2010). Additional research results expression that students still involvement problems in understanding the net force concept and concerning it to Newton's law (Hermanto et al., 2019; Lee & Park, 2013). Moreover, students also have misconceptions about fundamental concepts such as mass, acceleration and force (Fratiwi et al., 2017; Kaniawati et al., 2019; Liu & Fang, 2016). Misconceptions are impervious toward the novel concepts and further scientific (Eymur & Geban, 2017; Hermita et al., 2018; Yang & Sianturi, 2019). Even students holding misconceptions will throw away the novel concept established (Gurel et al., 2015; Stein & Galili, 2015; Taber, 2013). Consequently, it is problematic for them to entrench a scientific conception, at the end, that will eventually hamper the attainment of an occupied understanding of teaching supplies. Misconceptions should not be permissible and must rejected directly.

Therefore, a diagnostic test needed that can distinguish conceptions accurately and detect student misconceptions, such as the four-tier test. So that, the purpose of this study was to develop a multi-representation of tier instrument on Newton's laws, namely Multi-representation of Tier Instrument on Newtons' law (MOTION), using Rasch analysis.

II. METHODOLOGY

The ADDIE (Analyse, Design, Develop, Implement, and Evaluation) model used as the research method. In the Analyse stage, we analyse the data of misconceptions that often occur in Newton's Law material from previous studies. Besides, we also analyse the compatibility of the material with the current curriculum in Indonesia. At the Design stage, MOTION designed in a four-tier format, with three representations namely verbal, figure, and mathematical. During the Develop stage, MOTION was developed into a multi-representation in a four-tier format like the previous design. At the Implement stage, we spread MOTION to four groups of students through the Google form. Each group worked on nine questions for each sub-material Newton's first law, Newton's second law, Newton's third law, and the types of force. Finally, in the Evaluation stage, all students' answers were analysed based on the student's conception category, scored, then evaluated using Rasch analysis.

1) Participant

Participants in this study were students in one of the senior high schools in the district of Sukabumi, West Java. Participants numbered 92 students (41 male students and 51 female students, with an average age of 15-17 years) who were in K-11. Participants divided into four groups with the following distribution.

Table 1: Distribution of Participants

Group	*Male Students*	*Female Students*	*Total*
Group I (Newton's first law)	12	11	23
Group II (Newton's second law)	7	16	23
Group III (Newton's third law)	12	11	23
Group IV (type of forces)	10	13	23
Total	41	51	92

2) Instrument

The research instrument used in this study is named the Multi-representation of Tier Instruments on Newton's law (MOTION). The test developed from a standard test namely Force Concept Inventory (FCI), in the form of a four-tier test. As a pre-, post-, and delayed test requirement, instruments also developed in the form of multi-representations in the form of a verbal, figure and mathematic. Students do not realize the similarity of the questions they are working on if we used three representations. MOTION consisted of 36 questions consisting of Newton's first law, Newton's second law, Newton's third law and the types of force. Each sub material consists of nine questions with three representations.

3) Scoring

Before analysing the data using Rasch, the students' answers were first analysed based on the conception category, according to Table 2. After categorizing the conceptions, then each conception was given a score each. Sound Understanding (SU) given a score of '4' because students can answer all questions correctly and surely. Partial Understanding (PU) given a score of '3' because there are incorrect or unsure answers. False Positive (FP) given a score of '2' because there are indications that students have misconceptions as students are wrong in giving reasons. False

Negative (FN) given a score of '1' because students can only give reasons but wrong answers to problems. Alternative Conception (AC) given a score of '0' because students cannot answer all questions correctly. Finally, No Coding (NC) was no given a score because there was one tier that students did not answer. Rasch analysis can predict the part that is not filled in by students for NC's scoring.

Table 2: Students' Conception Criteria for Four-tier Test

Conception Criteria	*Tier 1*	*Tier 2*	*Tier 3*	*Tier 4*
Sound Understanding (SU)	Correct	Sure	Correct	Sure
Partial Understanding (PU)	Correct	Sure	Correct	Not sure
	Correct	Not sure	Correct	Sure
	Correct	Not sure	Correct	Not sure
	Correct	Sure	Incorrect	Not sure
	Correct	Not sure	Incorrect	Sure
	Correct	Not sure	Incorrect	Not sure
	Incorrect	Sure	Correct	Not sure
	Incorrect	Not sure	Correct	Sure
	Incorrect	Not sure	Correct	Not sure
	Incorrect	Sure	Incorrect	Not sure
	Incorrect	Not sure	Incorrect	Sure
	Incorrect	Not sure	Incorrect	Not sure
False Positive (FP)	Correct	Sure	Incorrect	Sure
False Negative (FN)	Incorrect	Sure	Correct	Sure
Alternative Conception (AC)	Incorrect	Sure	Incorrect	Sure
No Coding (NC)	If not filling one or more items (tier)			

4) Data Analysis

We used Rasch analysis as a data analysis. The data were organized through Microsoft Excel software and examined exhausting Winstep software version 4.4.5. The Rasch analysis converts the scores of items measured, which is ordinal data addicted to an interval scale called logarithm odds unit (logit) (Arsad et al., 2013; Park & Liu, 2019). Data analysis conducted included validity (unidimensionality and fit statistics) and reliability (Cronbach Alpha, item reliability and person reliability). The instrument had a good unidimensionality quantity when the index of raw variance was overhead the typical of 40% and index of unexplained variance 1st contrast less than 3 for eigen value and less than 15% for observed value (Adams, Sumintono, Mohamed, & Noor, 2018; Fisher, 2007).

The item and person fit statistics direct the amount toward which the data attained are appropriate, reliable and in agreement through the rudimentary measures, as well as charitable data about the quality of the measurement (Setiawan et al., 2018). There are numerous signs which are central equally for the persons and the items. Roughly of them are psychometric properties, such as outfit mean square (MNSQ), outfit Z-standardized (ZSTD), and Point measure correlation (PT Measure Corr.). The typical assessment initiates through detecting the value of outfit MNSQ, in which the value must be among the intervals of 0.5 and 1.5 (Setiawan et al., 2018). This income that it is appropriate for measurement. Uncertainty the value of MNSQ does not untruth in those intervals, it is essential to study the gotten outfit ZSTD value, which is hypothetical to be between a range of - 2.0 to + 2.0 (Park & Liu, 2019), the representative that the data have practical expectedness. Finally, the value of PT-Measure Corr. can also be seen, with an acceptance range between 0.4 to 0.85. The items would be in the 'overfit' criteria if all three criteria met, 'fit-2' if two of the three criteria met, 'fit-1' if one of the three criteria met and 'misfit' if all the criteria not met.

Cronbach Alpha coefficient rendering to the Rasch analysis defined the interaction between the persons and the items (Adams et al., 2018; Setiawan et al., 2018). Item reliability index designates that the steadiness of item and person reliability index designates that the constancy of person answers (Adams et al., 2018). If the value is near towards 1, it implies that the consistency of internal measurement is excellent.

III. Results And Analysis

The result sections will be described according to the ADDIE (Analyse, Design, Develop, Implement and Evaluation) model as follows.

Analyse

In the analyse stage, the data of misconceptions that often occur in Newton's Law material from previous studies analysed. Besides, we also analyse the compatibility of the material with the current curriculum in Indonesia, including Newton's first law, Newton's second law, Newton's third law and type of forces. The results of the analysis stage for the concepts and indicators presented in Table 3.

Table 3: The Result of the Analysis Stage

Concepts	*Indicators*
Inertia	Estimating the direction of motion of an object based on the principle of inertia
Force balance	Explain the force balance acting on stationary objects
	Explain the force balance acting on the regular straight motion
Effect of force on the velocity	Estimating the effect of force on the velocity of objects on horizontal motion
Effect of force on acceleration	Compare traces of falling objects based on the effect of mass on acceleration in free fall motion
Effect of motion on acceleration	Compare the effect of mass on acceleration in the inclined plane
Action-reaction force	Explain the action-reaction force on two colliding objects
	Explain the action-reaction force on a stationary object
	Grouping action-reaction forces on coincide objects
Normal force	Infer the normal force acting on objects in a moving elevator
Friction	Compares friction in the horizontal and inclined plane
Gravity	Describe the forces that work at the highest point

Design

At design stage, all of indicators at analyse stage were design as the four-tier test, as in Figure 2 below.

Tier 1
Question
…
…
…
…
…
Tier 2
Sure
Not sure
Tier 3

Reason … … … … … Tier 4 Sure Not Sure

Figure 2: Designing of Four-tier Test

The first tier in the four-tier test is an original multiple choice where problems presented, and then students asked to choose one correct answer from the five choices provided. The second tier is the confidence rating in the answers at the first tier. The third tier is the choice of reasons for the answer at the first tier, where students asked to choose one correct reason from the five choice reasons. Finally, the fourth tier is the confidence rating in the choice of reasons in the third tier. When the questions are verbal, the first and third tiers only presented with verbal explanations without mathematical symbols or figures. For questions in the form of figures, the first and third tiers are presented problems and answer choices in the form of pictures, graphs, or bar charts. Furthermore, for questions in the form of mathematics, the first and third tiers are presented in mathematical symbols, only slightly verbal and without figures.

Develop

At the developing stage, the instruments are developed based on indicators and designs. One indicator developed into three representations, namely verbal, figure and mathematic. The example of the instrument in verbal representation is:

3.1 An elevator is lifting an elevator shaft with steel cables. All friction effects ignored. When the elevator is moving up at a constant velocity, then ... A. the upward force by the cable is greater than the downward gravity force B. the upward force by the cable is the same as the downward force of gravity C. the upward force by the cable is smaller than the downward gravitational force D. the elevator rises because the cable is shorter, not because of the force applied by the cable to the elevator E. the upward force by the cable is greater than the downward force caused by air pressure and gravity 3.2 Are you sure about your answer to problem 3.1? A. Sure B. Not sure 3.3 Reasons for answers to questions 3.1: A. There is no acceleration of the elevator so that the upward force becomes larger. B. There is an acceleration of the elevator with an upward direction so that the upward force becomes greater. C. There is an acceleration of the elevator with an upward direction so that the upward force becomes smaller. D. There is no acceleration of the elevator, so the resultant force acting is also equal to zero. E. There is no velocity of the elevator so that the upward force becomes greater. 3.4 Are you sure about your answer to question 3.3? A. Sure B. Not sure

The example of the instrument in figure representation is:

3.1 An elevator is lifting an elevator shaft with a steel cable, as shown in Figure 3.

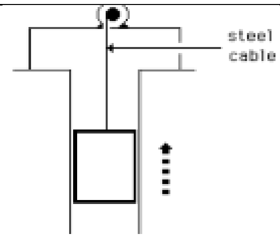
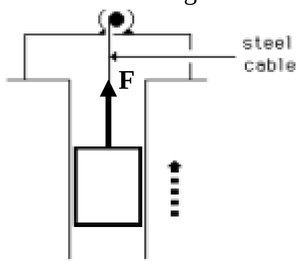


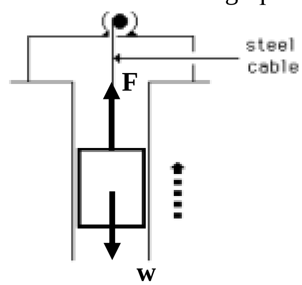
Figure 3. Shaft elevator that is moving up

All friction effects ignored. When the elevator is moving up at a constant velocity, then….

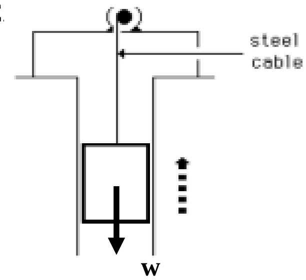
A.



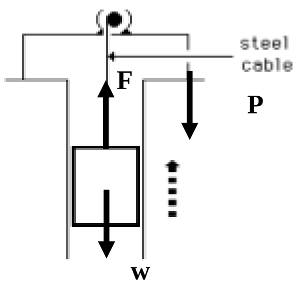
B.



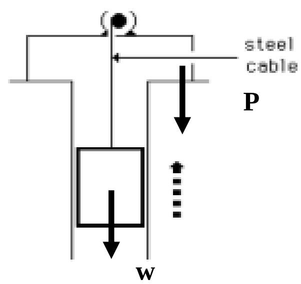
C.



D



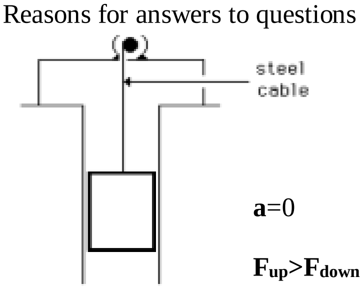
E.



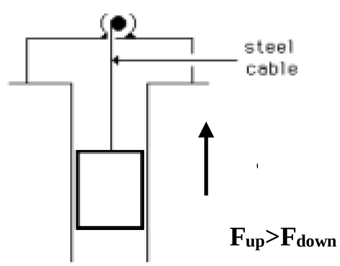
3.2 Are you sure about your answer to problem 3.1?
A. Sure
B. Not sure

3.3 Reasons for answers to questions 3.1:

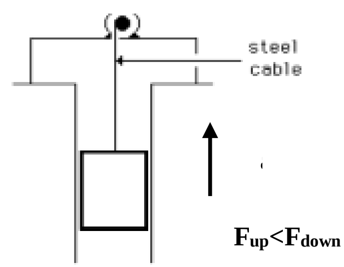
A.



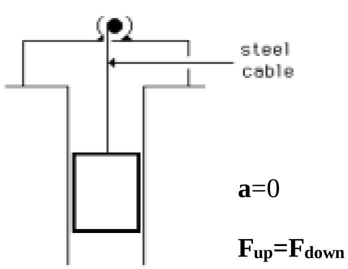
B.



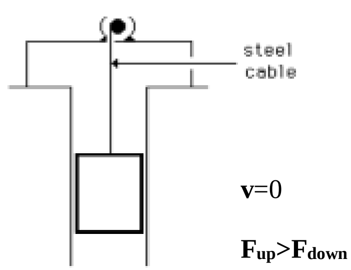
C.



D.



E.



3.4 Are you sure about your answer to question 3.3?
A. Sure
B. Not sure

The example of the instrument in mathematics representation is:

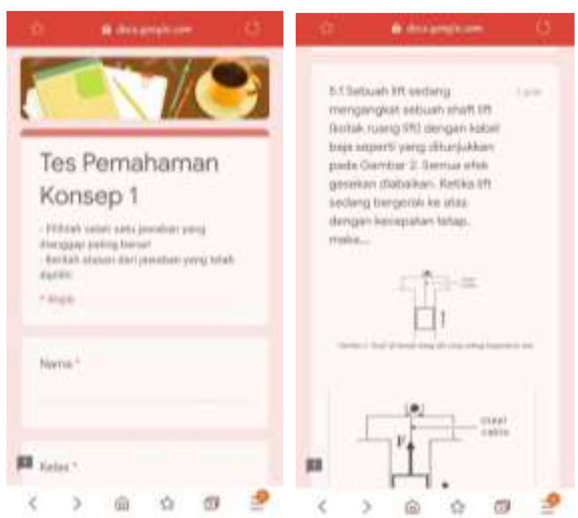
Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

ISSN: 1475-7192

3.1 An elevator is lifting an elevator shaft with steel cables. All friction effects ignored. When the elevator is moving up at a constant velocity, then ...
A. $\mathbf{F} > \mathbf{w}$
B. $\mathbf{F} = \mathbf{w}$
C. $\mathbf{F} < \mathbf{w}$
D. $\mathbf{F} > \mathbf{w}$ and $\mathbf{P}$
E. $\mathbf{F} = 0$, the elevator rises because the cable becomes shorter
3.2 Are you sure about your answer to problem 3.1?
A. Sure
B. Not sure
3.3 Reasons for answers to questions 3.1:
A. $\mathbf{a} = 0$ so $\mathbf{F}$ >>.
B. $\mathbf{a} = 0$ so $\mathbf{F}$ <<.
C. $\mathbf{a} \neq 0$ so $\mathbf{F}$ >>.
D. $\mathbf{a} \neq 0$ so $\mathbf{F}$ <<.
E. $\mathbf{a} = 0$ so $\Sigma\mathbf{F} = 0$.
3.4 Are you sure about your answer to question 3.3?
A. Sure
B. Not sure

Implement

During the implementing stage, MOTION distributed to four different groups. Each group gets three indicators with three representations so that the whole group does nine questions. The nine questions were randomized, so students did not realize they were working on the questions with the same indicators. The questions are distributed in the form of Google form, as shown in Figure 3, so students can work through laptops, computers or mobile phones.



(a) (b)

Figure 3: (a) Initial Display and (b) Question Display of Tests of Conceptual Understanding on Newton's First Law

Evaluation

At the evaluation stage, student responses at the implement stage are collected and then processed based on the conception criteria in Table 2. After the student's conception obtained, the conception then converted to scoring. Scoring data in Microsoft Excel then processed using Winstep software version 4.4.5. so that the validity and reliability of MOTION gained. The results of unidimensionality (validity) shown in Table 4 below.

Table 4: The Result of Unidimensionality based on Rasch Analysis

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

Sub Material	*Raw variance explained by measures*	*Interpretation*	*Unexplained variance 1st contrast*		*Interpretation*	*Decision*
			Eigen	*Observed*		
Newton's first law	64.9%	Excellent	2.0359	7.9%	Fulfilled	Valid
Newton's second law	51.3%	Good	2.6888	14.6%	Fulfilled	Valid
Newton's third law	52.0%	Good	2.8045	15%	Fulfilled	Valid
Type of forces	66.3%	Excellent	2.5675	9.6%	Fulfilled	Valid

From Table 4, if viewed from the raw variance explained by measures, for Newton's first law and the types of forces sub material are in the category of 'excellent' because the value is more than 60%, while for Newton's second law and Newton's third law sub material are in the category of 'good' because the value is more than 40%. Besides, if viewed from the eigen and observed values, all sub material is in the 'fulfilled' category because the eigen values are less than 3 and the observed value is less than 15%. Therefore, it can conclude that MOTION is valid for use. The processing of the fit statistic shown in Table 5 below.

Table 5: The Result of the Fit Statistic based on Rasch Analysis

Sub Material	*Question number*	*Outfit*		*PT Measure Corr*	*Interpretation*
		MNSQ	*ZFTD*		
Newton's first law	1 (verbal)	0.50	-1.02	0.83	Fit-2
	1 (figure)	2.38	2.05	0.63	Fit-1
	1 (mathematic)	1.42	0.92	0.59	Overfit
	2 (verbal)	0.84	-0.10	0.73	Overfit
	2 (figure)	0.63	-0.43	0.79	Overfit
	2 (mathematic)	1.08	0.33	0.69	Overfit
	3 (verbal)	0.49	-0.74	0.73	Overfit
	3 (figure)	0.75	-0.39	0.75	Overfit
	3 (mathematic)	1.76	1.42	0.60	Fit-2
Newton's second law	4 (verbal)	0.91	0.08	0.80	Overfit
	4 (figure)	3.01	2.31	0.67	Fit-1
	4 (mathematic)	1.02	0.26	0.82	Overfit
	5 (verbal)	0.81	0.20	0.93	Fit-2
	5 (figure)	0.29	-0.58	0.90	Fit-1
	5 (mathematic)	0.27	-0.77	0.92	Fit-1
	6 (verbal)	0.27	-0.84	0.90	Fit-1
	6 (figure)	0.94	0.25	0.86	Fit-2
	6 (mathematic)	0.25	-1.01	0.90	Fit-1
Newton's third law	10 (verbal)	1.41	0.92	0.68	Overfit
	10 (figure)	0.58	-0.92	0.76	Overfit
	10 (mathematic)	0.85	-0.02	0.60	Overfit
	11 (verbal)	1.04	0.25	0.73	Overfit
	11 (figure)	1.38	0.71	0.53	Overfit
	11 (mathematic)	1.63	1.21	0.70	Fit-2
	12 (verbal)	1.16	0.45	0.64	Overfit
	12 (figure)	0.34	-0.41	0.47	Fit-2
	12 (mathematic)	0.23	-1.30	0.69	Fit-2
Type of forces	7 (verbal)	0.93	0.04	0.65	Overfit
	7 (figure)	0.29	-1.66	0.87	Fit-1
	7 (mathematic)	0.71	-0.45	0.81	Overfit
	8 (verbal)	0.68	-0.52	0.82	Overfit
	8 (figure)	0.35	-1.66	0.84	Fit-2
	8 (mathematic)	3.36	3.14	0.65	Fit-1

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

Sub Material	*Question number*	*Outfit*		*PT Measure Corr*	*Interpretation*
		MNSQ	*ZFTD*		
	9 (verbal)	1.04	0.26	0.83	Overfit
	9 (figure)	1.01	0.20	0.78	Overfit
	9 (mathematic)	0.64	-0.54	0.83	Overfit

Based on Table 5, it can see that the items are in the overfit, fit-2, and fit-1 categories, and there are no misfit items. Overfit means that all criteria (MNSQ, ZSTD, PT Measure Corr.) met, fit-2 means that only two criteria met, and fit-1 means that only one criterion met. In Newton's first law there is one fit-1, in Newton's second law there are five fit-1 and in types of forces there are two fit-1. And then, the results of processing the reliability test are shown in Table 6 below.

Table 6: The Result of Cronbach Alpha, Item Reliability and Person Reliability based on Rasch Analysis

Sub Material	*Cronbach Alpha*	*Interpretation*	*Item Reliability*	*Interpretation*	*Person Reliability*	*Interpretation*	*Decision*
Newton's first law	0.89	Very good	0.85	Very good	0.80	Good	Reliable
Newton's second law	0.94	Very good	0.48	Poor	0.73	Good	Reliable with revision
Newton's third law	0.87	Very good	0.79	Good	0.48	Poor	Reliable
Type of forces	0.95	Very good	0.29	Poor	0.78	Good	Reliable with revision

From Table 6, it can see that the Cronbach Alpha value for all sub materials is in the 'very good' criteria because the value is more than 0.8. However, for item reliability, Newton's second law and types of forces have 'poor' reliability because their values are less than 0.5. In Newton's third law, person reliability is in the 'poor' category because its value is also less than 0.5, but for item reliability it is good. Therefore, Newton's second law and the types of forces are still reliable, but there is a need revision.

IV. Discussion And Conclusion

Multi-representation of Tier Instrument on Newtons' law (MOTION) has been developed based on the ADDIE model. Before developing a research instrument, it is necessary to analyse aspects that support the assembly of instruments such as material analysis based on the existing curriculum and previous research. Furthermore, the design of the instrument to be developed is needed, in this research, the four-tier test design. After analysis and design stage, instrument developed into a four-tier according to initial design that created. Besides, the instrument also developed into three representations, namely verbal, figure and mathematic. In order for the instrument to trust for use in larger research, the instrument needs to analyse for its validity and reliability. Thus, at the evaluation stage, the validity and reliability of the instrument were analysed using Rasch analysis. It is because the research instruments necessity to be valid and reliable concurrently (López-Lozano, Solís, & Azcárate, 2018; Summers, Wang, Abd-El-Khalick, & Said, 2019; Ursachi, Horodnic, & Zait, 2015). A measurement instrument is valid when it truly measures what it is hypothetical to measure (Ding, Wei, & Mollohan, 2016; Galili, 2019; Peter, 1981). From Table 4, it can be the decision that overall MOTION valid to use. This is in accordance with Adams et al. (2018) and Fisher (2007) research that the instrument had a good unidimensionality quantity when the index of raw variance was overhead the typical of 40% and index of unexplained variance 1st contrast less than 3 for eigen value and less than 15% for observed value. The validity of the instrument can also be seen for each item, whether there are misfit items or not. Based on Table 5, there are no misfit items, but some items are in the fit-1. Fit-1 means that the item can still measure what it should measure (valid), but it would be better if the item was revised.

Good instruments are not only valid but also reliable. Reliability is the scope to which an instrument will create consistent outcomes on comparable subjects below comparable situations and can be integrated with the correctness of a convinced measurement (Romine & Sadler, 2016; Tiruneh, De Cock, Weldeslassie, Elen, & Janssen, 2017; Ursachi et al., 2015). Reliability differs on how much variation in scores is attributable to chance faults (Zhu & Han, 2011). From Table 6, for Newton's second law and types of force, item reliability is in the 'poor' category. Because reliability is related to validity, it can see from the fit statistics (Table 5). For Newton's second law, five items are in the fit-1 category, and for types of force, two items are in the fit-1 category. This shows that reliability in the 'poor' category is due to the fact that there are several items in the fit-1 category. That is the item only meets one criterion out of three fit item criteria, so the item must be revised. Zhu & Han (2011) revealed some features that affect reliability, such as length and difficulty on the test paper itself. The longer questions always display more reliability than short ones. This is because the more substances there are in the test paper. If there are more characteristic substances in the test paper, the reliability of the test will be more complete. The level of difficulty will also affect testing reliability. If the questions in the test are either very difficult or very easy, the reliability of the test paper will equally be inclined.

Person reliability shows the consistency of the person on the item. In Table 6, the person reliability for Newton's third law is in the 'poor' category. This shows that there is a problem with the person factor. Zhu & Han (2011) and Al-Kalbani, Al Barwani & Neisler (2020) state that some factors such as sickness, emotions or motivation will restrict the person's ordinary testing level so that the consistency of a test is condensed. Although there is a 'poor' category for reliability, overall it can be concluded that MOTION is reliable for use. It is because a valid measurement is reliable, but a reliable measurement is not certainly valid (Galili, 2019; Ludwig, Priemer, & Lewalter, 2019; Ursachi et al., 2015). Eventually, other researchers can use MOTION to diagnose student conceptions. In addition, because this MOTION uses three representations, each representation can be used to evaluate the same student at different times. Moreover, Rasch analysis also can be used to develop a research instrument.

REFERENCES

[1] Adadan, E. (2013). Using multiple representations to promote grade 11 students' scientific understanding of the particle theory of matter. Research in Science Education, 43(3).

[2] Adadan, E., & Savasci, F. (2012). An analysis of 16-17-year-old students' understanding of solution chemistry concepts using a two-tier diagnostic instrument. International Journal of Science Education, 34(4), 513–544.

[3] 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(2), 227–256.

[4] Adimayuda, R., Aminudin, A. H., Kaniawati, I., Suhendi, E., Samsudin, A. (2020). A multitier open-ended momentum and impuls (MOMI) instrument: Developing and assesing quality of conception of 11th grade sundanese students with rasch analysis. International Journal of Scientific & Technology Research, 9(2), 4799-4804.

[5] Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. Learning and Instruction, 16, 183–198.

[6] Al-Kalbani, M., Al Barwani, T., & Neisler, O. (2020). Psychometric properties and factor structure of the university readiness survey. International Journal of Psychosocial Rehabilitation, 24(1), 1-8.
[7] Aminudin, A. H., Kaniawati, I., Suhendi, E., Samsudin, A., Coştu, B., & Adimayuda, R. (2019). Rasch Analysis of Multitier Open-ended Light-Wave Instrument (MOLWI): Developing and Assessing Second-Years Sundanese-Scholars Alternative Conceptions. Journal for the Education of Gifted Young Scientists, 7(3), 607–629.
[8] Arsad, N., Kamal, N., Ayob, A., Sarbani, N., Tsuey, C. S., Misran, N., & Husain, H. (2013). Rasch model analysis on the effectiveness of early evaluation questions as a benchmark for new students ability. International Education Studies, 6(6), 185–190.
[9] Boone, W. J., Abell, S. K., Volkmann, M. J., Arbaugh, F., & Lannin, J. K. (2011). Evaluating Selected Perceptions of Science and Mathematics Teachers in an Alternative Certification Program. International Journal of Science and Mathematics Education, 9, 551–569.
[10] Brass, C., Gunstone, R., & Fensham, P. (2003). Quality learning of physics: Conceptions held by high school and university teachers. Research in Science Education, 33(2), 245–271.
[11] Caleon, I., & 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.
[12] Chen, C. H. H. I. H., Lin, H. U. H., & Lin, M. I. N. G. I. (2003). Developing a Two-Tier Diagnostic Instrument to Assess High School Students ' Understanding − The Formation of Images by a Plane Mirror. Proceedings of the National Science Council, 106–121.
[13] Ding, L., Wei, X., & Mollohan, K. (2016). Does Higher Education Improve Student Scientific Reasoning Skills? International Journal of Science and Mathematics Education, 14(4), 619–634.
[14] Eymur, G., & Geban, Ö. (2017). The Collaboration of Cooperative Learning and Conceptual Change: Enhancing the Students' Understanding of Chemical Bonding Concepts. International Journal of Science and Mathematics Education, 15(5), 853–871.
[15] Fisher, W. P. (2007). Rating Scale Instrument Quality Criteria. Rasch Measurement Transactions, 21(1), 1095.
[16] Fratiwi, 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.
[17] Fratiwi, N. J., Utari, S., & Samsudin, A. (2019). Study of concept mastery of binocular K-11 students through the implementation of A multi-representative approach. International Journal of Scientific and Technology Research, 8(8), 1637–1642.
[18] Fratiwi, N. J., Samsudin, A., Ramalis, T. R., Saregar, A., Diani, R., Irwandani, I., Rasmitadila, & Ravanis, K. (2020). Developing MeMoRI on Newton's laws: for identifying students' mental models. European Journal of Educational Research, 9(2), 699-708.
[19] Galili, I. (2019). Towards a refined depiction of nature of science. Science & Education, 28(3–5), 503–537.
[20] Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. Eurasia Journal of Mathematics, Science and Technology Education, 11(5), 989–1008.
[21] Henke, A., & Höttecke, D. (2015). Physics teachers' challenges in using history and philosophy of science in teaching. Science and Education, 24, 349–385.
[22] Hermanto, I. M., Muslim, M., Samsudin, A., & Maknun, J. (2019). K-10 students ' conceptual understanding on Newton ' s laws : current and future directions. Journal of Physics: Conference Series, 1280, 1–6.
[23] Hermita, N., Suhandi, A., Syaodih, E., Samsudin, A., Mahbubah, K., Noviana, E., & Kurniaman, O. (2018). Constructing VMMSCCText for re-conceptualizing students ' conception. Journal of Applied Environmental and Biological Sciences, 8(3), 102–110.
[24] Hutajulu, M., Minarti, E. D., & Senjayawati, E. (2019). Improving of mathematical proficiency and disposition using multi representation approach on vocational students. Journal of Physics: Conference Series, 1–6.
[25] 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 and Technological Education, 35(2), 238–260.
[26] Kaniawati, I., Fratiwi, N. J., Danawan, A., Suyana, I., Samsudin, A., & Suhendi, E. (2019). Analyzing students' misconceptions about Newton's laws through four-tier Newtonian test (FTNT). Journal of Turkish Science Education, 16(1), 110–122.
[27] Kohl, P. B., Rosengrant, D., & Finkelstein, N. D. (2007). Strongly and weakly directed approaches to teaching multiple representation use in physics. Physical Review Special Topics - Physics Education Research, 3(010108).
[28] Krell, M., Redman, C., Mathesius, S., Krüger, D., & van Driel, J. (2018). Assessing Pre-Service Science Teachers' Scientific Reasoning Competencies. Research in Science Education.

[29] Kurnaz, M. A., & Arslan, A. S. (2014). Effectiveness of Multiple Representations for Learning Energy Concepts: Case of Turkey. Procedia - Social and Behavioral Sciences, 116, 627–632.
[30] Lamb, R. L., Annetta, L., Meldrum, J., & Vallett, D. (2012). Measuring Science Interest: Rasch Validation of The Science Interest Survey. International Journal of Science and Mathematics Education, 10, 643–668.
[31] Laszlo, P. (2013). Towards Teaching Chemistry as a Language. Science and Education, 22(7), 1669–1706.
[32] Lee, H. S., & Park, J. (2013). Deductive Reasoning to Teach Newton's Law of Motion. International Journal of Science and Mathematics Education, 11, 1391–1414.
[33] Liampa, V., Malandrakis, G. N., Papadopoulou, P., & Pnevmatikos, D. (2019). Development and Evaluation of a Three-Tier Diagnostic Test to Assess Undergraduate Primary Teachers' Understanding of Ecological Footprint. Research in Science Education, 49(3), 711–736.
[34] Liou, P. Y., & Hung, Y. C. (2015). Statistical Techniques Utilized in Aanalyzing PISA and TIMSS Data in Science Education from1996 to 2013: A Methodological Review. International Journal of Science and Mathematics Education, 13(6), 1449–1468.
[35] 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.
[36] López-Lozano, L., Solís, E., & Azcárate, P. (2018). Evolution of Ideas About Assessment in Science: Incidence of a Formative Process. Research in Science Education, 48(5), 915–937.
[37] Ludwig, T., Priemer, B., & Lewalter, D. (2019). Assessing Secondary School Students' Justifications for Supporting or Rejecting a Scientific Hypothesis in the Physics Lab. Research in Science Education, 1–26.
[38] Malone, K. L. (2008). Correlations among knowledge structures, force concept inventory, and problem-solving behaviors. Physical Review Special Topics - Physics Education Research, 4(020107).
[39] Murshed, M. B., Phang, F. A., Bunyamin, M. A. H. B., & Binti, I. J. (2020). The reliability analysis for force concept inventory. International Journal of Psychosocial Rehabilitation, 24(5), 143-151.
[40] Özcan, Ö., & Bezen, S. (2016). Students' mental models about the relationship between force and velo city concepts. Journal of Baltic Science Education, 15(5), 630–641.
[41] Panaoura, A., Michael-Chrysanthou, P., Gagatsis, A., Elia, I., & Philippou, A. (2017). A Structural Model Related to the Understanding of the Concept of Function: Definition and Problem Solving. International Journal of Science and Mathematics Education, 15, 723–740.
[42] Park, M., & Liu, X. (2019). An Investigation of Item Difficulties in Energy Aspects Across Biology, Chemistry, Environmental Science, and Physics. Research in Science Education.
[43] Perry, L. (2019). Development of an early grade relational reasoning subtask: collecting validity evidence on technical adequacy and reliability. International Journal of Science and Mathematics Education.
[44] Peşman, H., & Eryilmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. Journal of Educational Research, 103(3), 208–222.
[45] Peter, J. P. (1981). Construct validity: A review of basic issues and marketing practices. Journal of Marketing Research, 18(2), 133–145.
[46] Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.
[47] Romine, W. L., & Sadler, T. D. (2016). Measuring changes in interest in science and technology at the college level in response to two instructional interventions. Research in Science Education, 46, 309–327.
[48] Saglam-Arslan, A., & Devecioglu, Y. (2010). Student teachers' levels of understanding and model of understanding about Newton's laws of motion. Asia-Pacific Forum on Science Learning and Teaching, 11(1), 1–20.
[49] Samsudin, A., Azura, Kaniawati, I., Suhandi, A., Fratiwi, N. J., Supriyatman, Wibowo, F. C., Malik, A., & Costu, B. (2019). Unveiling students' misconceptions through computer simulation-based PDEODE learning strategy on dynamic electricity. Journal of Physics: Conference Series, 1280, 1-8.
[50] Samsudin, A., Azizah, N., Sasmita, D., Rasmitadila, Fatkhurrohman, M. A., Supriyatman, Wibowo, F. C. (2020). Analyzing the students' conceptual change on kinetic theory of gases as a learning effect through computer simulations-assisted conceptual change model. Universal Journal of Educational Research, 8(2), 425-437.
[51] Saputra, O., Setiawan, A., Rusdiana, D., & Muslim. (2020). Analysis of students' misconception using four tier diagnostic test on fluid topics. International Journal of Advanced Science and Technology, 29(1), 1256-1266.
[52] Setiawan, B., Panduwangi, M., & Sumintono, B. (2018). A Rasch analysis of the community's preference for different attributes of Islamic banks in Indonesia. International Journal of Social Economics, 45(12), 1647–1662.
[53] Sia, D. T., Treagust, D. F., & Chandrasegaran, A. L. (2012). High school students' proficiency and confidence levels in displaying their understanding of basic electrolysis concepts. International Journal of Science and Mathematics Education, 10, 1325–1345.

[54] Stein, H., & Galili, I. (2015). the Impact of an Operational Definition of the Weight Concept on Students' Understanding. International Journal of Science and Mathematics Education, 13(6), 1487–1515.
[55] Summers, R., Wang, S., Abd-El-Khalick, F., & Said, Z. (2019). Comparing Likert Scale Functionality Across Culturally and Linguistically Diverse Groups in Science Education Research: an Illustration Using Qatari Students' Responses to an Attitude Toward Science Survey. International Journal of Science and Mathematics Education, 17, 885–903.
[56] Sutopo, & Waldrip, B. (2014). Impact of a Representational Approach on Students' Reasoning and Conceptual Understanding in Learning Mechanics. International Journal of Science and Mathematics Education, 12, 741–766.
[57] Taber, K. S. (2013). Upper Secondary Students' Understanding of the Basic Physical Interactions in Analogous Atomic and Solar Systems. Research in Science Education, 43(4), 1377–1406.
[58] Tesio, L. (2003). Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. Journal of Rehabilitation Medicine, 35(3), 105–115.
[59] Theasy, Y., Wiyanto, & Sujarwata. (2018). Multi-representation ability of students on the problem solving physics. Journal of Physics: Conference Series, 1–4.
[60] Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. International Journal of Science and Mathematics Education, 15(4), 663–682.
[61] Treagust, D. (1986). Evaluating Students' Misconceptions by Means of Diagnostic Multiple Choice Items. Research in Science Education, 16, 199–207.
[62] Treagust, D. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. International Journal of Science Education, 10(2), 159–169.
[63] Ursachi, G., Horodnic, I. A., & Zait, A. (2015). How Reliable are Measurement Scales? External Factors with Indirect Influence on Reliability Estimators. Procedia Economics and Finance, 20(15), 679–686.
[64] Van Zile-Tamsen, C. (2017). Using rasch analysis to inform rating scale development. Research in Higher Education, 58(8), 922–933.
[65] Velentzas, A., & Halkia, K. (2013). From earth to heaven: Using "Newton's Cannon" thought experiment for teaching satellite physics. Science and Education, 22(10), 2621–2640.
[66] Yang, D. C., & Sianturi, I. A. J. (2019). Sixth grade students' performance, misconceptions, and confidence when judging the reasonableness of computational results. International Journal of Science and Mathematics Education, 17(8), 1519–1540.
[67] You, H. S., Marshall, J. A., & Delgado, C. (2018). Assessing students' disciplinary and interdisciplinary understanding of global carbon cycling. Journal of Research in Science Teaching, 1–25.
[68] Yuruk, N., Beeth, M. E., & Andersen, C. (2009). Analyzing the effect of metaconceptual teaching practices on students' understanding of force and motion concepts. Research in Science Education, 39, 449–475.
[69] Zhu, J., & Han, L. (2011). Analysis on the main factors affecting the reliability of test papers. Journal of Language Teaching and Research, 2(1), 236–238.