ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/347443581

Developing Multitier Instrument of Fluids Concepts (MIFO) to Measure Student's Conception: A Rasch Analysis Approach

Article *in* Journal of Advanced Research in Dynamical and Control Systems · October 2020

DOI: 10.5373/JARDCS/V12I6/S20201273

CITATIONS
0

READS
114

6 authors, including:



Nurdini Nurdini
Universitas Pendidikan Indonesia
6 PUBLICATIONS **3** CITATIONS
SEE PROFILE



Andi Suhandi
Universitas Pendidikan Indonesia
173 PUBLICATIONS **698** CITATIONS
SEE PROFILE



Taufik Ramalis
Universitas Pendidikan Indonesia
80 PUBLICATIONS **51** CITATIONS
SEE PROFILE



Achmad Samsudin
Universitas Pendidikan Indonesia
235 PUBLICATIONS **727** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Development HOT (Higher Order Thinking) Laboratory View project

K-10 Students' Conceptual Understanding of Linear Motion Concept: A Case Study View project

All content following this page was uploaded by Nurdini Nurdini on 18 December 2020.

The user has requested enhancement of the downloaded file.

Developing Multitier Instrument of Fluids Concepts (MIFO) to Measure Student's Conception: A Rasch Analysis Approach

Nurdini Nurdini[1], Andi Suhandi[1],Taufik Ramlan Ramalis[1], Achmad Samsudin[1], Nuzulira Janeusse Fratiwi[1], Bayram Costu[2]*
[1]*Department of Physics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia*
[2]*Department of Science Education, Yildiz Technical University, Turkey*

Abstract-This study aimed to develop a multitier instrument of fluids concepts (MIFO) using Rasch analysis. The research method as utilized the ADDIE model. Participants were 103 students (47male and 56female), and their average age was 16-17 years old, at one of senior high school in Bandung, West Java, Indonesia.The MIFO consists of 12 items in form of four-tier test. MIFO examined through Rasch analysis based onitem-person maps of student's conception, unidimensionality, fit statistic, item-person separation, reliability,and differential item functioning. Student's conceptions were categorized as scientific conception, lack of knowledge, and misconception. The outcomes of a Rasch analysis revealed how the evaluation analysis offered dataintended for assessingthe student's conception of fluids concepts. Then it can be concluded that the MIFO can be developed to measure student's conception. This study can be a reference for educators or teachers forfurther study and prepare tests to identify student's conceptions.

Keywords: Fluid concepts, Multitier Instrument of Fluids Concept (MIFO), Rasch analysis, student's conception

Introduction

The multitier instrument is one of the diagnostic tests that capable tool to identify student's conception, which aquestion combined with confidence levelsis a major effort to generating consciousness. However, the consciousness of being correct forms an essential component of one's knowledge and capability to understand[1], [2]. The multitier instrument permits educators to separate students who have scientific conception (i.e., confidence-accurate), who have a lack of knowledge (i.e., confidence-inaccurate), and who have a possible misconception (i.e., confidence-inaccurate). It has been suitable in several formats, as well as the four-tier format, which contains a discreterating of confidence for each of the two tiers [1], [3], [4], [5]. The diagnostic tests were first developed by Treagust [6]in the arrangement of a two-tier structure. In the first tier, students answer the question, and students give reasons for the selected answers of the first tier in the second tier. However, the two-tier format is challenging to differentiate amongprecise understanding or speculation[1], [7], [8], [9], [10]. As for three-tier tests still similarly have severalrestrictions due to the concealed rating of the confidence for the first and second tiers in those tests [11], [12], [13]. Therefore, the four-tier test can be more accurate to measure student's conception because one tier is added about the rating of confidence student's answer, so that the rating of confidence given does not become ambiguous for the first tier or the second tier[12], [14], [15], [16], [17], [18]. Developing the multitier instrument may offer insight intostudentswho have scientific conception and differentiate instancesthrough a potential lack of knowledge from those who clasp robust misconceptions[19]. The result is valuable advice that can be used by equally students and their educators to enhance learning and teaching, singly[20].

In physics concept, one of the misconceptions that are often experienced by students in the field of mechanics is the concept of fluid[21], [22]. Several studies mention that misconceptions in the concept of fluid, including hydrostatic pressure, depending on the height of the liquid above the size of the vessel[23], the buoyancy force of an object is proportional to the volume above the liquid surface[14], [24], an object will sink in liquid if its density is higher than the density of the liquid[24], light and small objects will float, and substantial, and large objects will sink[14], [25]. Specifically, for identifying the state of conception mentioned earlier, a multitier instrument in the concept of fluid was developed. The multitier instrument called MIFO (Multitier Instrument of Fluid Concepts).In outline, the flow of MIFO development can be seen in Figure 1.

DOI: 10.5373/JARDCS/V12I6/S20201273
**Corresponding Author: Nurdini Nurdini, Email: nurdini@upi.edu*
Article History: Received: Aug 02, 2020, Accepted: Oct 04, 2020

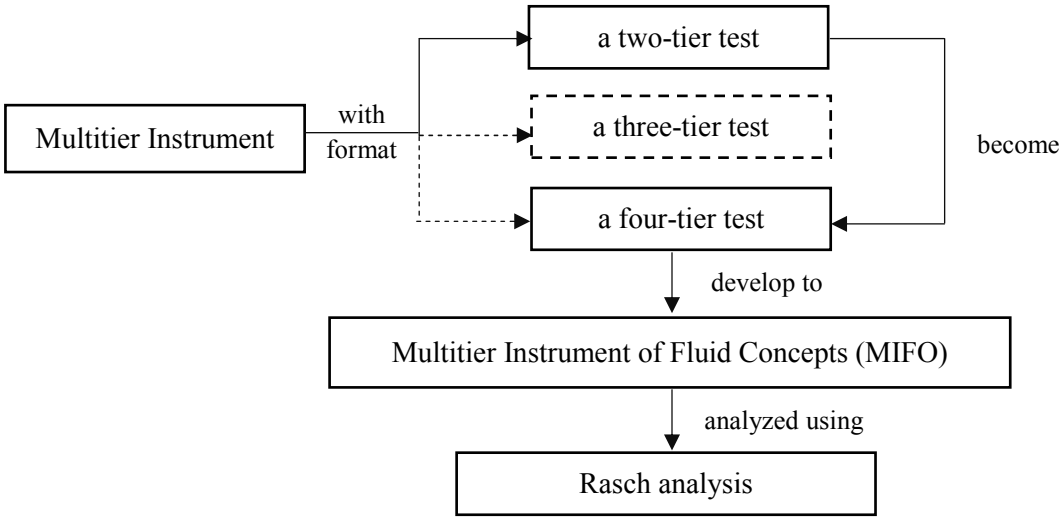


Figure 1 Development of the MIFO

The MIFO design in the four-tier test format for measuringstudent's conceptions. The first tier of the MIFO is an unadventurous multiple-choice test. The second tier is the confidence rating of the response in the first tier. The third tier requires the reasoning for the answer in the first tier, and the fourthtier involves the confidence rating of the answer in the third tier. This format is the same as the four-tier format in general based on several research instruments development in other concepts[12], [13], [14], [22], [23], [24]. A concept instrument must acquireexcellentpsychometric things[29]. However, the majority of instruments are analyzed by conventional practice, which is still dominantin makingutilization of validity along withbasic statistical procedurestowardevaluate its effectivenessmeant for diagnostic analysis[30], [31]. Whereas this is a permissible exercise, the accessibility of other instrumentspermits for anadditionalforcefulevaluation of the psychometric properties of the instrument, such as using Rasch Analysis. Several researchescorrelated to the validity of instruments with Rasch analysis of which have been conducted byKreijns, Bijker, & Weidlich[32], Nielsen & Santiago[33], You, Haudek, Merrill, & Urban-Lurain [34].

Rasch analysis can currently be used towardassess the process of developing a test and is a powerful psychometric tool [35], [36], [37], [38]. The quality of test measurements can be evaluated to revise test instruments to improve measurement precision based on the test data are collected.The Rasch analysis is a stochastic version that consents raw test elements to be exposed to a linear change to produce a standardized score that detects student capability and item difficulty on a standard logit scale[39]. Rasch analysis can be used for cases of dichotomous or polytomous. In this study, the data obtained from the MIFO were polytomous because the data analyzed were in the form of student's conceptions. However, the polytomous response structure may be furtherappropriate in variouscircumstances. The polytomous Rasch analysis for ordered response categories [29], [33], [34], [35], [36], [37]can be labeled mathematically through the subsequent equation:

$$\Pr\{x_{ni} = x\} = \frac{e^{-\tau_{1i}-\tau_{2i}\cdots-\tau_{xi}+x(\beta_n-\delta_i)}}{\sum_{x'=0}^{m_i} e^{-\tau_{1i}-\tau_{2i}\cdots-\tau_{x'i}+x'(\beta_n-\delta_i)}}$$

Therefore, this study aims to develop multitier instruments on the concept of fluid, which is named MIFO and will be analyzed differently using Rasch analysis. The development of instruments examined by Rasch analysis, especially in developing instruments on the concept of fluid, is significant because it is still rarely done. Thus, the MIFO developed can be used accurately to identify the category of student's conceptions.

Method

Research Design

The research method used an ADDIE (Analyzing, Designing, Developing, Implementing, and Evaluating) model. The analyzing phase has been completed to literature researchesregardingmisconceptionson fluid concepts and tools for analyzing instrument tests. The designingphase is carried out through designing a multitier test instrument with a four-tier format. Furthermore, at the instrument developing stage, a multitier instrument to measure student's conceptions of the fluid concept was developed, which was then called the MIFO. Then the MIFO is tested on students during the implementing stage. The last is the evaluating stage to ascertainwhether this

instrument is suitable for use in measuring student's conceptions. In this stage, the instrument was evaluated using Rasch analysis based on each student's conception category from the results of the implementingstage.

Participants

The participants(N = 103) arestudents of 11th-grade classes in one of the senior high schools in Bandung, West Java, Indonesia. The participants were randomly selected from the two-gender (i.e., males and females), with most participants aged around 16-17 years. The female students included 56 students of the total participants, and the remaining 47 student examinees were male students. All participants are students who have learned the concept of fluid.

Data Analysis

The data analyzed were the acquisition of each student's conception category according to the conception criteria of Kaltakci-Gurel, Eryilmaz, &McDermott[12]. Then each category of student conception is given a score (i.e., students whose scientific conception has a score of 2, lack of knowledge has a score of 1, and misconceptions have a score of 0) so that it can be analyzed using Rasch analysis. Rasch analysis was completed with Winsteps software version 4.4.7.

Item analysis includes the level of difficulty of the items carried out as an initial investigation of the test items indicated in item person maps. Furthermore, unidimensionality is the underlying assumption in Rasch analysis. The unidimensionality assumption involves that there is only one fundamental construct restrained by the set of items involved in the test is that the test measures only one factor. Instrument unidimensionality is an important measure to evaluate whether the instrument being developed can measure what should be restrained so that it can be said to be valid. Rasch analysis uses principal component analysis of standardized residual variance (in Eigenvalue units) [30], [35], [45], [46].A validity test based on the instrument's unidimensionality can be seen in the value of raw variance explained by measures. Regarding this, the criteria for unidimensionality, which includes the criteria "fulfilled"if the amount of variance explained by measures be>20%, if the value>40% means "good", and if the value >60 means "excellent". Alsoto find out whether or not there are problem items and misfit, it can be seen in the eigenvalue value and observed in unexplained variance 1st contrast with the eigenvalue value must be less than <3.0 to indicate there are no problem items and the observed value must be <15% to indicate item fit.

The criteria used in analyzing the suitability of these items can be seen in fit order items by reviewing the outfit values of MNSQ, ZSTD, and PT Measure Corr.[47], [48]. The criteria for each are (0.5 <x <1.5) for MNSQ, (-2 <x <+2) for ZSTD, and (0.4 <x <0.85) for PT Measure Corr. If all three criteria are met, then it can be said the item "overfit" can be ascertained that the quality of the item is good to use. It says "fit" if one or two criteria are met then the item can still be maintained and need not be changed, whereas "misfit "If the three criteria are not met then it can be ascertained that the items are not good enough so that they need to be fixed or replaced.

Furthermore, item-person separation and reliability with the Rasch analysis are based on the interfaceamong the person and the entire item that can be shown based on the Cronbach Alpha value. Besides that, it is also based on the consistency of student's responsesfollowing the level of student's abilities called person reliability. It is based on the quality of the items in the instrument that can be shown based on the value of item reliability. Related to the value of separation, if the separation value is more significant, then the quality of the instrument in terms of the overall of respondents and item show better. For items, fit can indicate the quality of an item in performing its function as a measuring tool.

Meanwhile, the detection of bias for items in the Rasch analysis is displayed in the differential item function (DIF). DIF is only necessary to find out whether the items given have biases in specific categories of respondents or not. Bias in items can be known based on the probability value of items that are under 5%. In general, the flow of data analysis can be shown as follows:

DOI: 10.5373/JARDCS/V12I6/S20201273
*Corresponding Author: Nurdini Nurdini, Email: nurdini@upi.edu
Article History: Received: Aug 02, 2020, Accepted: Oct 04, 2020

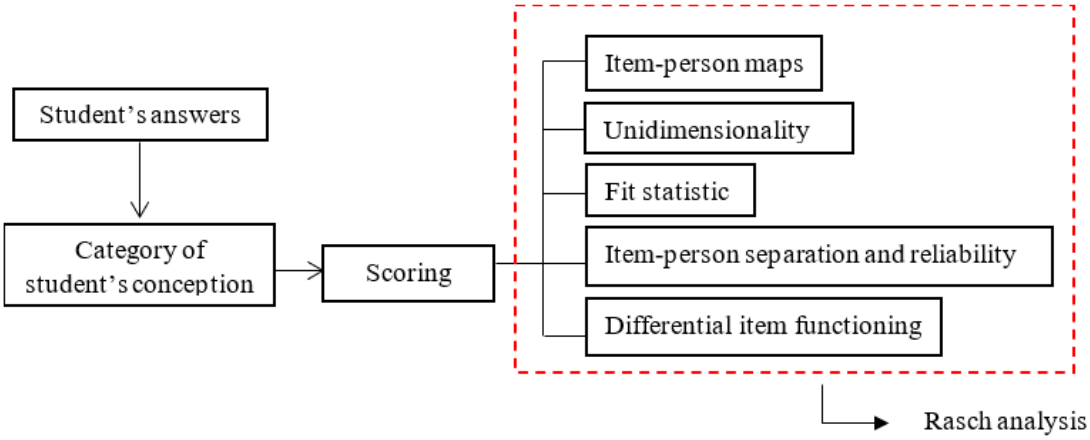


Figure 2 Flowchart of research analysis

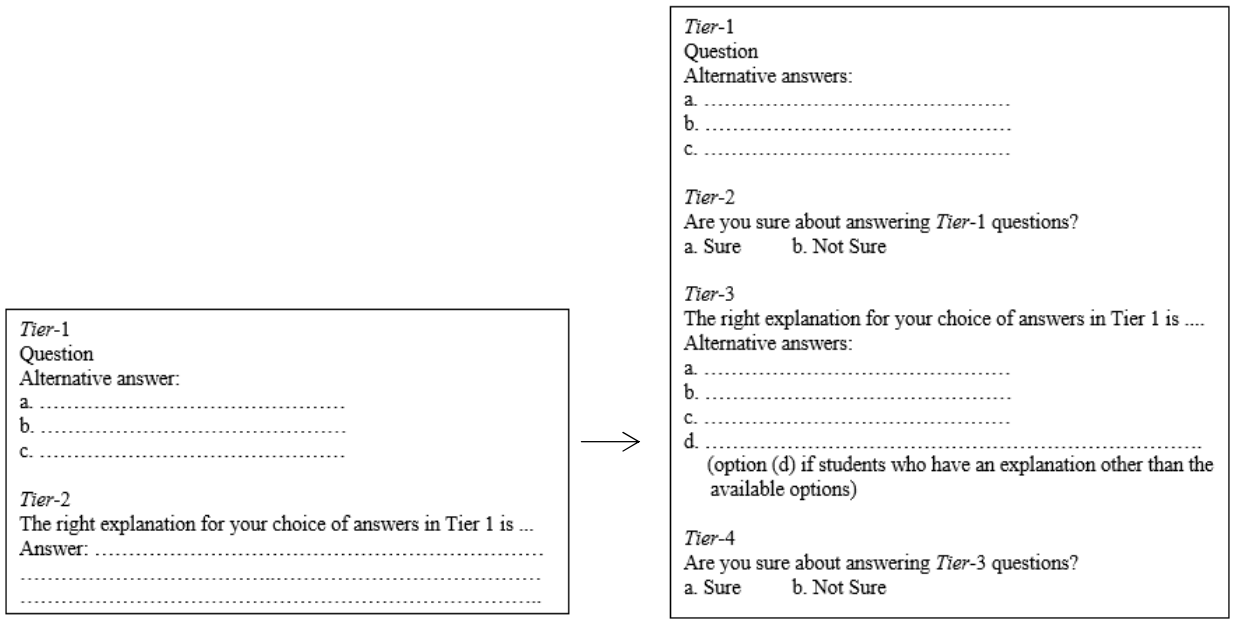
Results

Analyzing

At the analyzing phase, researchers examined the literature study about developing instruments of student's conceptions, misconceptions on fluid concepts, and tools for analyzing instrument tests. The results of the analysis of several studies indicate that some instrument development has also been carried out on physics concepts such as Force Concept Inventory [49], Mechanics Baseline Test [50], Conceptual Survey of Electricity & Magnetism [51], and Wave Diagnostic Instrument [5], Multitier Open-ended Light-Wave Instrument (MOLWI) [52], Multi-representation of Tier Instrument on Newton's law(MOTION) [13], etc. However, the analysis of the instruments used is also diverse, and dominant still uses conventional practice. In this connection, the researcher chooses the concept of fluid to be analyzed in terms of the kinds of misconceptions encountered by students and uses instrument analysis with Rasch analysis. The results of the misconception study analysis show that three fluid sub-concepts often occur in student's misconceptions, including the concept of hydrostatic pressure, the concept of buoyancy force, and the concepts of floating, drifting, and sinking.

Designing

Furthermore, at the designing stage, the design used on the MIFO in measuring student's conceptions from open to closed questions as done by Aminudin et al. [52], where open-ended is a two-tier test. The MIFOis designed with a four-tier format. The conception test in a four-tier format consists of the first tier in the form of multiple-choice answers. The second tier is the rating of student's confidence in the choice of answers at the first tier.The third tier is the reason for the answers at the first level. Finally, the fourth tier is the rating of confidence learners towards the choice of reasons at the third tier. The four-tier design from the two-tier open-ended into a four-tier close-ended instrument shown in Figure 3.



(a) (b)

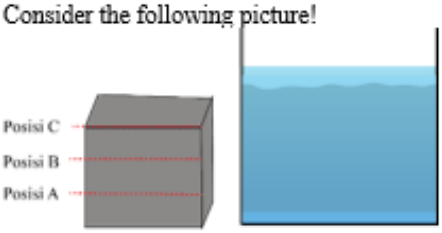
Figure 3(a) Two-tier design, (b) Four-tier design

Developing

At the developing stage, the instruments are developed based on four-tier design. The MIFO contains 12 items, including the concept of hydrostatic pressure, the concept of buoyancy, and the concepts of floating, drifting, and sinking, each of the three items. The choice of reasons at the third tier in the four-tier test presented is the result obtained from a student's answers when conducting a study of the two-tier test. In other words, the reasons given by students are analyzed to be alternative answers at the third tier in the four tiers. Examples of MIFO that have been created are shown in Figure 4.

DOI: 10.5373/JARDCS/V12I6/S20201273
**Corresponding Author: Nurdini Nurdini, Email: nurdini@upi.edu*
Article History: Received: Aug 02, 2020, Accepted: Oct 04, 2020

Tier-1
Consider the following picture!



A brick is hung with a stative when dipped in water in various positions. At position A the brick is dipped 1/3 of the volume of the brick, at position B the brick is dipped 2/3 of the volume of the brick, and position C the brick is immersed entirely. The great buoyancy style of bricks on bricks A, B, and C is ...
a. the brick buoyancy in positions A, B, and C is the same.
b. the buoyant force of the brick when dipped to position A is greater than position B and C.
c. the buoyant force of the brick when dipped to position C is greater than position A and B.

Tier-2
Are you sure about answering *Tier*-1 questions?
a. Sure b. Not Sure

Tier-3
The right explanation for your choice of answers in Tier 1 is ....
Alternative answers:
a. the buoyant force depends on the volume of the liquid, because the conditions of the liquid volume at position A, B, and C are equal, then the buoyancy force of the bricks at each position is equal.
b. the buoyant force is proportional to the volume of the object on the surface of the liquid, the greater the volume of the object on the surface of the liquid shows the greater the buoyancy of the object and vice versa. In position A, the volume of the brick is above the water level so that in position A the buoyancy force of the brick is greater than in other positions.
c. buoyancy is proportional to the volume of objects immersed in liquid, when the volume of objects immersed in liquid is greater than the volume of objects on the surface of liquid shows that the buoyancy of objects is greater and vice versa. At position C, the volume of the bricks that are immersed is greater so that the buoyancy force at position C is greater than other positions.

Tier-4
Are you sure about answering *Tier*-3 questions?
a. Sure b. Not Sure

Figure 4 The Example of the MIFO

Implementing

Then in the implementation stages, researchers distributed the MIFO to 103 students online using the web, so MIFO can be accessed via smartphones, computers, or laptops by students.Students can access MIFO on http://physicedu.epizy.com/ as shown in Figure 5. Students fill in the items on the MIFO for 45 minutes.

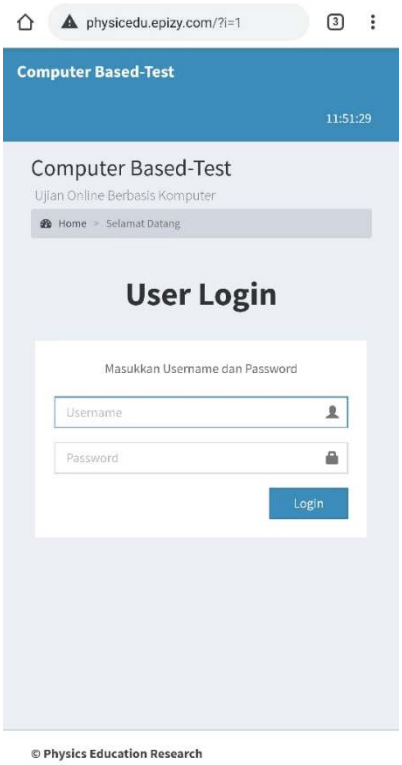


Figure 5 MIFO display on the web

Evaluating

The last, at the evaluating stage, student's answers from the implementation results are collected and analyzed based on the conception criteria. Furthermore, scoring data using Microsoft Excel and analyzed with Winstep software version 4.4.7 to evaluate the MIFO that have been developed. The MIFO evaluation carried out covering item-person maps of student's conception, unidimensionality, fit statistic, item-person separation and reliability, and differential item functioning. Item-person maps of student's conception outcome to indicate the level of difficulty items can be shown in Figure 6.

DOI: 10.5373/JARDCS/V12I6/S20201273
*Corresponding Author: Nurdini Nurdini, Email: nurdini@upi.edu
Article History: Received: Aug 02, 2020, Accepted: Oct 04, 2020

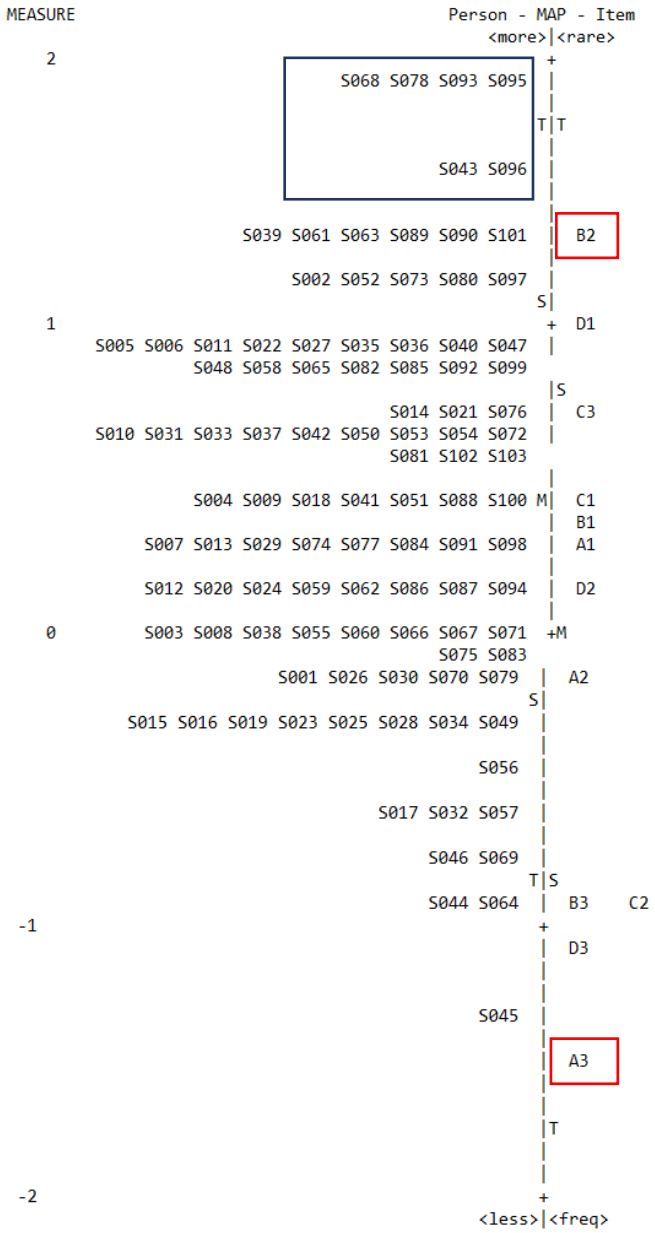


Figure 6 The result of item-person of student'sconception

Figure 6 shows the results of the variable maps of items and persons. On the right side shows the map item, while the left side shows the map person (student). Students on the top are students who have high ability, while the lowest has a low ability. The top item (B2) position is the most challenging item to answer for students, while the item at the bottom (D3) is the easier item for students to answer.Based on the person map, the students at the top are students who have high abilities, while students who are ranked are students with low abilities. Also, there were six students (S068, S078, S093, S095, S043, and S096) of 103 students whose abilities could not be accommodated by the 12 items developed. So, it is necessary to have items added to accommodate the measurement of the six students' abilities.

Furthermore, the results of unidimensionality as construct validity can be seen in Table 1.

Table 1 The result of unidimensionality

Concepts	Item Number	Raw Variance Explained by Measures	Interpretation	Unexplained Variance 1st Contrast		Interpretation
				Eigenvalue	Observed	

Concepts	Item Number	Raw Variance Explained by Measures	Interpretation	Unexplained Variance 1st Contrast		Interpretation
				Eigenvalue	Observed	
Hydrostatic Pressure	A1, A2, A3	65.4%	Excellent	1.4489	*16.7%	Fulfilled
Buoyancy Force	B1, B2, B3	78.8%	Excellent	1.7367	12.3%	Fulfilled
Floating, Drifting, and Sinking	C1, C2, C3	59.8%	Good	1.4244	*19.1%	Fulfilled
	D1, D2, D3	81.4%	Excellent	1.7762	11.0%	Fulfilled

The sign (*) in the Unexplained variance 1st contrast: observed indicates the need to review item fit order

From Table 1, the results of the raw variance value explained by measure show that the items on the concept of hydrostatic pressure (A1, A2, A3) and the concept of buoyancy force (B1, B2, B3), and the concepts of floating, drifting, and sinking (D1, D2, D3) is in the "excellent" category.While for the floating, drifting, and sinking concept in items C1, C2, and C3 are included in the "good" category. Furthermore, based on the eigenvalue value on the unexplained variance 1st contrast shows that there are no problem items because the value obtained is less than 3. Then, based on the acquisition of the observed value shows the items on the concept of buoyancy (B1, B2, B3). For floating, drifting, and sinking concepts (D1, D2, D3) shows that the six items are suitable for use.While the items on the concept of hydrostatic pressure (A1, A2, A3) and the concepts of floating, drifting and sinking (C1, C2, C3) obtain an observed value of more than 15, which indicates a tendency to misfit items. So that further analysis can be done with the analysis of item fit order to determine whether an item can be maintained or must be replaced. Based on the findings of eigenvalue and observed, the items in the three concepts can be said to be "fulfilled".The findings of the item fit are shown in Table 2.

Table 2 The findings of item fit

Concepts	Item Number	Outfit		PT-Measure Corr.	Interpretation
		MNSQ	ZFTD		
Hydrostatic Pressure	A1	1.72	1.55	0.79	Item Fit
	A2	0.88	-0.36	0.82	Item Fit
	A3	2.91	1.56	0.55	Item Fit
Buoyancy Force	B1	0.61	-1.83	0.81	Item Fit
	B2	0.78	-0.31	0.64	Item Fit
	B3	5.84	2.61	0.68	Item Fit
Floating, Drifting, and Sinking	C1	0.93	-0.25	0.79	Item Fit
	C2	2.20	1.86	0.57	Item Fit
	C3	0.94	-0.09	0.77	Item Fit
	D1	1.75	1.37	0.76	Item Fit
	D2	0.40	-3.91	0.84	Item Fit
	D3	1.82	1.05	0.59	Item Fit

For the record, the ZSTD value is strongly affected by the sample. When the sample is very large, certainly, the ZSTD value will always be above 3.From Table 2 shows that there are five items (A2, B1, B2, C1, C3) that meet all three criteria, five items (A2, A3, C2, D1, D3) that meet two criteria, and two items (B3, D2) that meet one criterion. Nevertheless, overall items can be categorized as fit. Then, instrument reliability test results based on item reliability and person reliability, along with the separation results, are shown in Table 3 below.

Table 3 The result of item-person separation and reliability

Concepts	Item Number	Person Reliability	Separation	Interpretation	Item Reliability	Separation	Interpretation

DOI: 10.5373/JARDCS/V12I6/S20201273
*Corresponding Author: Nurdini Nurdini, Email: nurdini@upi.edu
Article History: Received: Aug 02, 2020, Accepted: Oct 04, 2020

Concepts	Item Number	Person Reliability	Separation	Interpretation	Item Reliability	Separation	Interpretation
Hydrostatic Pressure	A1, A2, A3	0.82	0.71	Good	0.97	5.44	Excellent
Buoyancy Force	B1, B2, B3	0.63	1.32	Sufficient	0.88	7.92	Good
Floating, Drifting and Sinking	C1, C2, C3	0.37	0.77	Weak	0.97	6.21	Excellent
	D1, D2, D3	0.89	1.48	Good	0.99	8.50	Excellent

Based on Table 3shows that the consistency of answers from students can be said to be "good" with the quality of items in the "excellent" instrument on the concept of hydrostatic pressure (A1, A2, A3) and floating, drifting, and sinking (D1, D2, D3). On the concept of buoyancy style (B1, B2, B3), the consistency of students is "sufficient" with the quality of the items in the "good" instrument. In contrast, the concept of floating, drifting, and sinking (C1, C2, C3) illustrates that the consistent answers from participants students "weak," but the quality of items in the instrument "excellent".

Besides, the reliability test results based on Cronbach Alpha values in measuring interactions between person and item are shown in Table 4.

Table 4 The result of reliabilitybased on Cronbach Alpha values

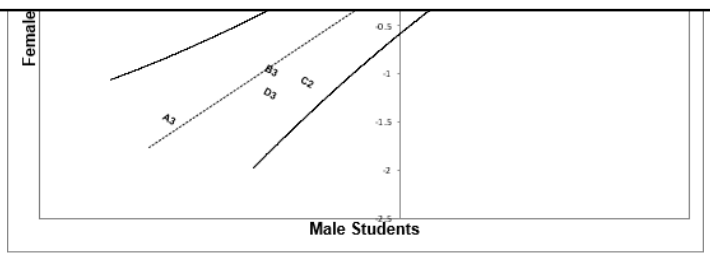
Concepts	Item Number	Cronbach Alpha Values	Interpretation
Hydrostatic Pressure	A1, A2, A3	0.67	Sufficient
Buoyancy Force	B1, B2, B3	0.72	Good
Floating, Drifting and Sinking	C1, C2, C3	0.83	Very Good
	D1, D2, D3	0.61	Sufficient

In Table 4, overall, based on the Cronbach Alpha value shown was acceptable. For differential item functioning (DIF) results in knowing which items can be shown in Figure 7. In the DIF analysis, the evaluation group is towards the gender group (male and female).

Figure 7 The result of differential item functioning

In Figure7, it can be perceived that each item does not contain any DIF, where the probability value for each

Person CLASSES	SUMMARY DIF CHI-SQUARED	D.F.	PROB.	BETWEEN-CLASS/GROUP Item UNWTD MNSQ	ZSTD	Number	Name
2	5.5074	1	.0589	5.7999	2.16	1	A1
2	3.4769	1	.0622	3.6111	1.60	2	A2
2	.0870	1	.7680	.0880	-.71	3	A3
2	.5661	1	.4518	.5752	.11	4	B1
2	.3027	1	.5822	.3079	-.22	5	B2
2	.0375	1	.8464	.0383	-.94	6	B3
2	.0000	1	1.0000	.0020	-1.38	7	C1
2	1.5981	1	.2062	1.6416	.85	8	C2
2	1.0662	1	.3018	1.0860	.53	9	C3
2	.6241	1	.4295	.6319	.17	10	D1
2	3.4729	1	.0624	3.5995	1.60	11	D2
2	.7061	1	.4007	.7192	.25	12	D3



item is above 5% (0.05). Also, the DIF results can be seen in Figure 8 with a scatter plot display.

Figure 8 The result of differential item functioning on the scatter plot to display

Based on Figure 8, there are no items that show off the line, so this also shows that there are no items that are biased towards a group of female students or male students. In the DIF analysis, the DIF detection curve can also be used to analyze items specifically for the level of difficulty of the problem with groups of female students and male students.The DIF detection curve can be seen in Figure 9.

Figure 9 The result of DIF detection curve

From Figure 9, there are three curves based on student gender (female students (F) and male students (M)), and the (*) sign indicates the average value of both. In Figure 7, it can be realized that compared to male students, the curves in items A1, A2, A3, C3, B1, and C3 for female students are relatively more challenging to do. However, for items B1, B2, C2, D1, D2, and D3, female students are better able to work on items than male students. For items of



B3 and C1, both groups (female students and male students) have the same ability to work on the items.The difficulty level of the questions, the turtle in Figure 7 shows that item B2 is the most challenging item, while B6 on the lowest curve is considered the most effortless item. The DIF detection curve result is consistent with the description related to item-person maps in Figure 6 before.

Discussion And Conclusion

The research used the ADDIE model to develop the Multitier Instrument of Fluid Concepts (MIFO). Also, Rasch analysis has been used to analyze the psychometric properties of the MIFO. In developing the instrument, the researcher has analyzed a literature review to find out various previous studies related to the development of student's conception instruments test, misconceptions on fluid concepts, and tools for analyzing instrument tests. It can be used as a guide to developing better instruments and with the evaluation process of different instruments using Rasch analysis. Subsequently,the MIFOis designed in a four-tier format following Figure 2.

Furthermore, the developingphaseis carried out to produce and arrange the MIFOfollowing the design so that it becomes a complete the MIFO in a four-tier format. To be able to evaluate more deeply, the MIFO needs to be implemented to students. The implementation stage is carried out for 103 students who are among the students of the 11th grade. The data obtained from the implementation stage is used as a material to be evaluated using Rasch analysis.

Generally, the instrument presentsrelatively good psychometric things. Research instruments must be valid and reliable at the same time [13], [53], [54]. Concerningthe validity, the criticalproblem is whether it senses to determine what it is intended to measure. The item-person map indicatesmoderately good pursuing though it also suggests the necessity to involve more complicatedelements to improveaccommodate to students of higher capability levels. This is because the range of people is around 1.80 logit and -1.40 logit while the range of items is approximately 1.30 logit and -1.60 logit.

DOI: 10.5373/JARDCS/V12I6/S20201273
*Corresponding Author: Nurdini Nurdini, Email: nurdini@upi.edu
Article History: Received: Aug 02, 2020, Accepted: Oct 04, 2020

It should be noted that the order of the level of difficulty of the results of item-person maps is the same as the results of the DIF detection curve. Dependable through studies from the item map, this out come additional recommends that the range of test items may not match the scale of ability levels that exist. Also, according to research from Chan & Subramaniam [30], the addition of more delicate items can also be done to increase the validity of the test as well as its ability to distinguish between people consistently.

Concern about the uni dimensionality of the MIFO suggests that calibration of the test of Rasch analysis revealed that there are many items misfit but based on the results of raw variance explained by measures show in the category of excellent and good. While in the eigenvalue and observed results, several items indicate a tendency to misfit items, so it needs to be further analyzed with item fit order analysis to determine whether an item can be maintained or must be replaced. However, with revisions to someitems, the MIFO was validly used to measuring student's conceptions. Relating the explained unidimensionality, the results of the data-model fit revealed 12 items were in acceptable range value of criteria item fit, with the tendency of items to fulfill one or two criteria. However, overall, the items are fit. The result of item fit shows that 12 items can be used as research instruments, and there are no items to be replaced.

From the results of reliable item-person, it can be shown that 9 of the 12 items are reliable. For the other three items, namely C1, C2, and C3 on the concept of floating, drifting, and sinking, can be said to be reliable, but there needs to be a revision. Whereas, based on the item-person separation, value shows a variety of values. However, overall the quality of the instrument in terms of items is good because it can identify item groups and person groups[55], [56]. In this regard, overall, the interaction between a person (students) and items is in the category of sufficient, good, and very good.

Then on the DIF for gender, which is principally apparent in student's conception scale, also needs attention. It's because possible gender could negatively impact one or more groups in a construct-irrelevant manner [45]. According tothe study ofCamilli [57],DIF analysis focuses on the performance of two or more different groups in this study analyzed by the gender group. For the most part, DIF for gender is not unexpected and can achieve a balance between items that female students and male students[58]. From the DIF results, it is interpretable that 12 items suggest that results of the MIFO test scores are free of construct-irrelevant variance. Hence, it does support the argument for construct validity. In line with the evaluation of the MIFO varied, DIF analysis across gender groups was investigated too. DIF analysis excluded a similar probability of ratifying the test items across the gender groups. The results indicated that all itemsdo not contain DIF. This is because of the probability value of all items above 5% [45]. It is meansrevealing that the results of the present study are consistent with Karami's[59]study from the aspect of unidimensionality, in which the multidimensionality in the whole test and among sub-tests had been proven.

Therefore, based on the discussion exposure, it can be concluded that the MIFO can be developed to measure student's conception using Rasch analysis. The significant contribution of this study in the field of developing instruments of student's conception is providing empirical evidence in the form of the MIFO that passes a comprehensive evaluation process using Rasch analysis.

Acknowledgments

The authors would like to thank all those who assisted in completing this research, especially to our friends and staff of the physics education department lecturer, who always provided useful support and discussion in this research.

References

[1] M. Versteeg, M. Wijnen-Meijer, and P. Steendijk, “Informing the uninformed: A multitier approach to uncover students’ misconceptions on cardiovascular physiology,” *Adv. Physiol. Educ.*, 2019.

[2] J. M. Tuma *et al.*, “Recognizing Uncertainties in Evaluating Responses,” *Ecosyst. Hum. Well-being Policy Responses*, 2005.

[3] Y. K. Yan and R. Subramaniam, “Using a multi-tier diagnostic test to explore the nature of students’ alternative conceptions on reaction kinetics,” *Chem. Educ. Res. Pract.*, 2018.

[4] B. Sreenivasulu and R. Subramaniam, “Exploring Undergraduates’ Understanding of Transition Metals Chemistry with the use of Cognitive and Confidence Measures,” *Res. Sci. Educ.*, 2014.

[5] I. S. Caleon and R. Subramaniam, “Do students know What they know and what they don’t know? Using a four-tier diagnostic test to assess the nature of students’ alternative conceptions,” *Res. Sci. Educ.*, vol. 40, no. 3, pp. 313–337, 2010.

[6] D. Treagust, “Development and use of diagnostic tests to evaluate students’ misconceptions in science,” *Int. J. Sci. Educ.*, vol. 10, no. 2, pp. 159–169, 1988.

[7] C. Y. Tsui and D. Treagust, “Evaluating secondary students’ scientific reasoning in genetics using a two-tier

diagnostic instrument," *Int. J. Sci. Educ.*, vol. 32, no. 8, pp. 1073–1098, 2010.
[8] H. O. Arslan, C. Cigdemoglu, and C. Moseley, "A Three-Tier Diagnostic Test to Assess Pre-Service Teachers' Misconceptions about Global Warming, Greenhouse Effect, Ozone Layer Depletion, and Acid Rain," *Int. J. Sci. Educ.*, vol. 34, no. 11, pp. 1667–1686, 2012.
[9] H. Peşman and A. Eryilmaz, "Development of a three-tier test to assess misconceptions about simple electric circuits," *J. Educ. Res.*, vol. 103, no. 3, pp. 208–222, 2010.
[10] U. Kanli, "A study on identifying the misconceptions of pre-service and in-service teachers about basic astronomy concepts," *Eurasia J. Math. Sci. Technol. Educ.*, 2014.
[11] D. K. Gurel, A. Eryilmaz, and L. C. McDermott, "A review and comparison of diagnostic instruments to identify students' misconceptions in science," *Eurasia J. Math. Sci. Technol. Educ.*, vol. 11, no. 5, pp. 989–1008, 2015.
[12] D. Kaltakci-Gurel, A. Eryilmaz, and L. C. McDermott, "Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics," *Res. Sci. Technol. Educ.*, vol. 35, no. 2, pp. 238–260, 2017.
[13] A. Samsudin, N. J. Fratiwi, T. R. Ramalis, A. H. Aminudin, B. Costu, and M. Nurtanto, "Using rasch analysis to develop multi-representation of tier instrument on newton's law (motion)," *Int. J. Psychosoc. Rehabil.*, 2020.
[14] N. Nurdini, T. R. Ramalis, and A. Samsudin, "Exploring K-11 Students ' Conception using a Four-Tier Diagnostic Test on Static Fluid : a Case Study," in *RSU Conference*, 2019, pp. 672–681.
[15] Rahmawati, N. Y. Rustaman, I. Hamidah, and D. Rusdiana, "The development and validation of conceptual knowledge test to evaluate conceptual knowledge of physics prospective teachers on electricity and magnetism topic," *J. Pendidik. IPA Indones.*, vol. 7, no. 4, pp. 483–490, 2018.
[16] R. Adimayuda, A. H. Aminudin, I. Kaniawati, E. Suhendi, and A. Samsudin, "A multitier open-ended momentum and impulse (MOMI) instrument: Developing and assessing quality of conception of 11th grade sundanese students with rasch analysis," *Int. J. Sci. Technol. Res.*, 2020.
[17] M. G. Purwanto, R. Nurliani, I. Kaniawati, and A. Samsudin, "Promoting the hydrostatic conceptual change test (HCCT) with four-tier diagnostic test item," *J. Phys. Conf. Ser.*, vol. 1013, pp. 1–6, 2018.
[18] V. Liampa, G. N. Malandrakis, P. Papadopoulou, and D. Pnevmatikos, "Development and Evaluation of a Three-Tier Diagnostic Test to Assess Undergraduate Primary Teachers' Understanding of Ecological Footprint," *Res. Sci. Educ.*, vol. 49, no. 3, pp. 711–736, 2019.
[19] I. Hussein, E. O. C. Mkpojiogu, and A. Hussain, "A focus group assessment of participants' UX knowledge and their motivations for participating in a UXD community of practice," *J. Adv. Res. Dyn. Control Syst.*, vol. 11, no. 05-Special Issue, pp. 1496–1499, 2019.
[20] J. Jamaluddin, G. Gunawan, M. Muhlis, P. Sedijani, and Y. A. S. Anwar, "Model of project based learning with computer simulations to improve conceptual understanding of prospective science teachers," *J. Adv. Res. Dyn. Control Syst.*, vol. 11, no. 10, pp. 180–185, 2019.
[21] A. Samsudin, G. D. Putra, D. Saepuzaman, A. H. Aminudin, and A. Rais, "A Development of PhET based Mechanical Fluid Worksheet to identify changes in student conceptions," *Test Eng. Manag.*, vol. 83, pp. 15441–15451, 2020.
[22] A. Samsudin *et al.*, "Improving students' conceptions on fluid dynamics through peer teaching model with PDEODE (PTM-PDEODE)," in *Journal of Physics: Conference Series*, 2018.
[23] U. Besson, "Students' conceptions of fluids," *Int. J. Sci. Educ.*, 2004.
[24] S. A. Kiray, F. Aktan, H. Kaynar, S. Kilinc, and T. Gorkemli, "A descriptive study of pre-service science teachers' misconceptions about sinking-floating," *Asia-Pacific Forum Sci. Learn. Teach.*, 2015.
[25] Y. Yin, M. Tomita, and R. Shavelson, "Diagnosing and Dealing with Student Misconceptions: Floating and Sinking.," *Sci. Scope*, 2008.
[26] N. J. Fratiwi, I. Kaniawati, E. Suhendi, I. Suyana, and A. Samsudin, "The transformation of two-tier test into four-tier test on Newton's laws concepts," in *AIP Conference Proceedings*, 2017.
[27] I. Kaniawati, N. J. Fratiwi, A. Danawan, I. Suyana, A. Samsudin, and E. Suhendi, "Analyzing students' misconceptions about Newton's Laws through Four-Tier Newtonian Test (FTNT)," *J. Turkish Sci. Educ.*, vol. 16, no. 1, pp. 110–122, 2019.
[28] O. Saputra, A. Setiawan, D. Rusdiana, and Muslim, "Analysis of students' misconception using four tier diagnostic test on fluid topics," *Int. J. Adv. Sci. Technol.*, 2020.
[29] H. Lee, H. Yang, Y. Lee, A. Lim, and Y. Song, "Psychometric properties of the Korean version of barriers

DOI: 10.5373/JARDCS/V12I6/S20201273
*Corresponding Author: Nurdini Nurdini, Email: nurdini@upi.edu
Article History: Received: Aug 02, 2020, Accepted: Oct 04, 2020

to insulin treatment scale," *J. Adv. Res. Dyn. Control Syst.*, no. 10-Special Issue, pp. 205–210, 2017.
[30] M. Chan and R. Subramaniam, "Validation of a science concept inventory by rasch analysis," in *Rasch Measurement: Applications in Quantitative Educational Research*, 2020.
[31] A. Stammen, "The Development and Validation of the Middle School-Life Science Concept Inventory (MS-LSCI) Using Rasch Analysis," 2018.
[32] K. Kreijns, M. Bijker, and J. Weidlich, "A rasch analysis approach to the development and validation of a social presence measure," in *Rasch Measurement: Applications in Quantitative Educational Research*, 2020.
[33] T. Nielsen and P. H. R. Santiago, "Using graphical loglinear rasch models to investigate the construct validity of the perceived stress scale," in *Rasch Measurement: Applications in Quantitative Educational Research*, 2020.
[34] H. S. You, K. Haudek, J. Merrill, and M. Urban-Lurain, "Construct validity of computer scored constructed response items in undergraduate introductory biology courses," in *Rasch Measurement: Applications in Quantitative Educational Research*, 2020.
[35] C. H. Yu, "Objective measurement: How rasch modeling can simplify and enhance your assessment," in *Rasch Measurement: Applications in Quantitative Educational Research*, 2020.
[36] F. Franchignoni, F. Horak, M. Godi, A. Nardone, and A. Giordano, "Using psychometric techniques to improve the balance evaluation systems test: The mini-bestest," *J. Rehabil. Med.*, 2010.
[37] S. S. Ringo and A. Samsudin, "Rasch Analysis : Measuring Students ‟ Attitudes toward Physics using the CLASS," *Test Eng. Manag.*, vol. 83, pp. 15461–15467, 2020.
[38] N. J. Fratiwi *et al.*, "Developing memori on Newton's laws: For identifying students' mental models," *Eur. J. Educ. Res.*, vol. 9, no. 2, pp. 699–708, 2020.
[39] W. J. Boone, "Rasch basics for the novice," in *Rasch Measurement: Applications in Quantitative Educational Research*, 2020.
[40] D. Andrich, "A rating formulation for ordered response categories," *Psychometrika*, 1978.
[41] U. Böckenholt and T. Meiser, "Response style analysis with threshold and multi-process IRT models: A review and tutorial," *Br. J. Math. Stat. Psychol.*, 2017.
[42] C. H. Yu, S. O. Popp, S. Digangi, and A. Jannasch-Pennell, "Assessing unidimensionality: A comparison of Rasch modeling, Parallel analysis, and TETRAD," *Pract. Assessment, Res. Eval.*, 2007.
[43] D. Andrich, "The Polytomous Rasch Model and the Equating of Two Instruments," in *Rasch Models in Health*, 2013.
[44] M. Hopman-Rock, S. Van Buuren, and M. De Kleijn-De Vrankrijker, "Polytomous Rasch analysis as a tool for revision of the severity of disability code of the ICIDH," *Disabil. Rehabil.*, 2000.
[45] S. M. Alavi and S. Bordbar, "Differential Item Functioning Analysis of High-Stakes Test in Terms of Gender: A Rasch Model Approach," *Malaysian Online J. Educ. Sci.*, 2016.
[46] J. Chow and A. Shiu, "Examining an economics test to inform university student learning using the rasch model," in *Rasch Measurement: Applications in Quantitative Educational Research*, 2020.
[47] W. J. Boone, J. R. Staver, M. S. Yale, W. J. Boone, J. R. Staver, and M. S. Yale, "Person Reliability, Item Reliability, and More," in *Rasch Analysis in the Human Sciences*, 2014.
[48] T. G. Bond and C. M. Fox, *Applying the Rasch Model. In Applying the Rasch model: Fundamental Measurement in the Human Sciences*. 2015.
[49] D. Hestenes, M. Wells, and G. Swackhamer, "Force concept inventory," *Phys. Teach.*, 1992.
[50] D. Hestenes and M. Wells, "A mechanics baseline test," *Phys. Teach.*, 1992.
[51] M. Planinic, "Assessment of difficulties of some conceptual areas from electricity and magnetism using the Conceptual Survey of Electricity and Magnetism," *Am. J. Phys.*, 2006.
[52] A. H. Aminudin, I. Kaniawati, E. Suhendi, A. Samsudin, B. Coştu, and R. Adimayuda, "Rasch analysis of Multitier Open-ended Light-Wave Instrument (MOLWI): Developing and assessing second-years sundanese-scholars alternative conceptions," *J. Educ. Gift. Young Sci.*, vol. 7, no. 3, pp. 607–629, 2019.
[53] R. Summers, S. Wang, F. Abd-El-Khalick, and Z. Said, "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," *Int. J. Sci. Math. Educ.*, vol. 17, pp. 885–903, 2019.
[54] L. Ding, X. Wei, and K. Mollohan, "Does Higher Education Improve Student Scientific Reasoning Skills?," *Int. J. Sci. Math. Educ.*, vol. 14, no. 4, pp. 619–634, 2016.
[55] B. Sumintono, "Rasch Model Measurements as Tools in Assesment for Learning," 2018.
[56] V. Mešić *et al.*, "Measuring students' conceptual understanding of wave optics: A Rasch modeling approach," *Phys. Rev. Phys. Educ. Res.*, vol. 15, no. 1, pp. 1–20, 2019.

[57] G. Camilli, "Test fairness," *Educ. Meas.*, vol. 4, pp. 221–256, 2006.
[58] M. Curtin, J. Browne, A. Staines, and I. J. Perry, "The Early Development Instrument: An evaluation of its five domains using Rasch analysis," *BMC Pediatr.*, 2016.
[59] H. Karami, "The quest for fairness in language testing," *Educ. Res. Eval.*, 2013.

DOI: 10.5373/JARDCS/V12I6/S20201273
**Corresponding Author: Nurdini Nurdini, Email: nurdini@upi.edu*
Article History: Received: Aug 02, 2020, Accepted: Oct 04, 2020

View publication stats