JIPF (JURNAL ILMU PENDIDIKAN FISIKA)
p-ISSN: 2477-5959 | e-ISSN: 2477-8451 *Vol. 8 No. 2, May 2023, Page 226-239*

Using Rasch Model to Analyze Reliability and Validity of Concept Mastery Test on Electricity and Magnetism Topic

Yusri Handayani [1], Rahmawati [2*)], Widiasih [3]
Universitas Muhammadiyah Makassar, Indonesia[1,2], Universitas Terbuka, Indonesia[3]
[*)]Corresponding E-mail: rahmawatisyam@unismuh.ac.id

Received: December 10th, 2022. Revised: January 15th, 2023. Accepted: January 31st, 2023

Keywords :
Concept Mastery Test; Rasch Model; Validity; Reliability; Winsteps 3.68.2

ABSTRACT
This study aims to analyze the reliability and validity of the electricity and magnetismconcept mastery test (EM-CMT) using Rasch model aided by Winsteps version 3.68.2. The multiple choice concept mastery test consists of 40 items about material electric current, resistance, resistivity, Ohm's law, electric voltage, energy and electrical conductivity, resistors in series, parallel, and mixed circuits, Kirchhoff's law, RC charging and discharging capacitor circuits, magnetic fields, magnetic forces, and sources magnetic field. The sample of this study were prospective physics teacher students who had passed basic physics courses. The results showed that the EM-CMThad a value of the Alpha Cronbach reliability with a good and acceptable category (0.70). The value of person reliabilityis in the quite good category (0.69). Meanwhile, the personnel separation coefficient of 1.50 which indicates the response of respondents is quite good and consistent. Analysis of the observational aspects of item fit shows that there are no items that need to be changed or removed. For the aspect of person fit, the results of the analysis showed that there were seventeen out of sixty-three respondents who experienced unusual response patterns. A review of the observations of the map variable shows the distribution of respondents' abilities and the items are proportional. Analysis of unidimensionalaspects shows the value of "raw variance explained by measure" is in the acceptable category. Based on the results of the analysis on a number of aspects, it can be concluded that theEM-CMT test is reliable and valid so that it can be used to measure the students' concepts.

INTRODUCTION

The electricity and magnetism materialsare one of the material groups from the distribution of material groups in basic physics courses. The abstract nature of electricity and magnetism makes this material group difficult for students to understand starting from elementary level education [1] [2] [3] [4] [5], intermediate level [6] [7] [8] [9] [10] [11] [12], college level [13] [14] [15], to the level of instructors (teachers) [6] [16] [17].

Electrical and magnetic material is an important fundamental concept and must be mastered by prospective physics teacher students to be able to take courses to a higher level, namely electric magnet courses and other courses with the characteristics of physics expertise content courses with different levels. higher. In addition, physics teacher candidates must master this material because they will teach the material at the elementary and secondary levels of education with a lower level of difficulty. In relation to student life, electrical and magnetic material is important to be mastered by prospective physics teacher students because the interaction of electricity and magnetism plays a central role in determining the structure of life and becomes the foundation of important technological developments in life [18] [19]. Therefore, the material of electricity and magnetism is important to be provided to prospective teacher students through lectures and learning assessments. To determine the level of mastery of the concept of physics teacher candidates related to electricity and magnetism, it is necessary to measure.

One technique that can be used in measuring the level of mastery of students' concepts is the test technique. In this case, the test is a set of items that are relevant to the purpose of the test and developed by the test maker [20] [21]. Various types of test forms that can be used to measure the level of achievement or mastery of student concepts in lectures. One of them is an objective test form [22] [23] [24] [25].

The use of objective test forms is not limited to measuring learning outcomes that are simple. However, this form of test can be used to measure various learning outcomes knowledge [20] [26]. One of the most common and widely used types of objective tests in educational assessments is multiple choice tests [26] [27]. When compared with other types of tests, multiple choice tests can provide more adequate measurements [26]. In addition, this type of test can measure a variety of learning outcomes ranging from simple to complex levels and is easy to use and process the results of the tests [23] [26] [27]. However, the design of multiple choice test types is more complex and requires more time in preparation than other forms of test.

A test is called to be good if it fulfills three characteristics, namely valid, reliable, and usable so that it is able to accurately measure the ability of students [20] [21] [25] [28] [29]. A number of approaches to test the validation of test instruments, namely content validation, constructs, and criteria validation [20] [30]. In this regard, this study focuses on analyzing the level of reliability and validity of the instrument mastery test on the concepts of electricity and magnetism.

The concepts of validity and reliability are two very important concepts in the development and use of learning outcomes test instruments. Validity refers to the extent to which the evidence and theory support the interpretation of the scores required by users of the proposed test [20] [26] [27] [31]. Furthermore, validity is basically a collection of evidence to provide a scientific basis for interpretation of test scores [32]. Thus, it can be concluded that validity shows the ability of the test to measure what should be measured.

The quality of the validity of the test instrument can be seen from various types of evidence of the validity carried out. Sources of evidence of the validity of the test instruments used by the American Educational and Psychological Research Association (AERA and APA) can be divided into five types, namely: (1) proof of validity based on the contents of the test; (2) proof of validity based on the response process; (3) proof of validity based on internal structure; (4) proof of validity based on relationships with other variables; and (5) proof of validity based on testing consequences [31].

The second concept of test instrument validation is reliability. Reliability refers to the consistency of measurements, in this case, how the results of the test scores or the results of test tests are consistent from one measurement to the next that is done on a group of individuals or groups [20] [31]. Reliability is not only related to test results, but more than that, reliability is related to gathering enough information to ensure that the test reliability process represents the behavior of certain groups of students [29].

The process of test instrument validation in this study focuses on the empirical validation review of the validity of the internal structure and the level of test reliability. There are two types of approaches that can be done to analyze items related to the level of validity and reliability, namely the classical test theory approach and item response theory (IRT).

Classical test theory is based on an additive model, where the observed score is the sum of the actual score and measurement error score [33]. The classical test theory model is based on weak assumptions, where assumptions can be easily found through a set of test data and the model can be applied to a variety of test development and test score analysis problems [34] [35]. Classical test theory has a number of fundamental weaknesses, namely: (1) the classical test theory model uses a number of statistics such as the level of difficulty and differing items depending on the respondents tested in the analysis; (2) classical test theory is more oriented to the test results obtained than the items in the test instrument itself; (3) the concept of test reliability in the context of classical theory based on the alignment of test kits is very difficult to fulfill. In practice, it is very difficult to get two completely parallel test sets; (4) classical test theory does not provide a basis for determining how the respondents of test takers respond if given certain points; (5) the standard error index of measurement is assumed to be the same for each test taker; (6) the testing procedure for item bias and test equivalency are not practical and difficult to carry out. Likewise, equalization is vertical [33] [35]. For this reason, psychometrics offer an alternative measurement theory and model called the item response theory (IRT).

The item response theory model (IRT) shows the relationship between the ability or nature measured by the instrument and an item response [34] [36]. There are three assumptions underlying the item response theory, namely unidimensional, local independence, and parameter invariance. Unidimensional means that each test item only measures one ability. In practice, unidimensional assumptions cannot be met strictly because of external factors that are not constant. Local independence is a condition where external factors that influence constant achievement that causes the subject's response to any item will be statistically independent of each other. The assumption of local independence will be fulfilled if the participant's answer to an item does not affect the participant's answer to another item. Parameter invariance means that the characteristics of the items do not depend on the distribution of the test taker's ability parameters and the parameters that characterize the test takers do not depend on the item characteristics [34] [35] [36].

A model that is popular in the use of item response theory (IRT) is known as the logistical model. There are three types of logistic models, namely one-parameter logistic model, two parameters of logical model, and three parameters of logistic models. The logical one-parameter model is one of the most widely used IRT models. This model is also called the Rasch Model. The mathematical form of the logical one-parameter model is given by equation (1) [34] [35].

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \quad (1)$$

The Rasch model was first introduced by George Rasch so that the one parameter logistic IRT application model was enshrined as his name [37] [38] [39]. The Rasch model is a psychometric technique developed to improve the precision of instruments that are constructed, to monitor instrument quality and to calculate respondent performance [40] [41]. The application of the Rasch model has been widely used in various related research fields: 1) learning of science and mathematics [42] [43] [44] [45] [46] [47]; 2) assessment of science and mathematics learning [32] [48] [49] [50] [51] [52]; 3) health [53] [54] [55]; 4) language [56]; 5) psychology [57].

The Rasch Model Analysis can provide information to users about the reliability of people and items, person and item separations, and the Cronbach Alpha value. Meanwhile, the construct validity of the test instrument can be assessed through item fit, wright maps, and unidimensional. Based on the criteria for determining the validity of the test instrument, this study will use these key concepts in analyzing the reliability and validity of the concept mastery test instrument developed using the help of the *Winsteps* program.

METHOD

This research is a research that uses survey methods from the results of the test instrument trials that have been developed. The analysis technique uses quantitative descriptive which is oriented to descriptive statistical analysis to analyse the reliability and validity of the test instrument that will be used to measure the level of mastery of the concept of prospective physics teacher students in the matter of electricity and magnetism.

The test Instrument of electrical and magnetic concept mastery (EM-CMT) in the form of multiple choice tests consists of 40 items developed from nine topics, namely electric current, resistance, resistivity, Ohm's law, electric voltage, energy and electrical conductivity, resistors in series, parallel, and mixed circuits, Kirchhoff's law, RC charging and discharging capacitor circuits, magnetic fields, magnetic forces, and sources magnetic field. This test instrument aims to measure the concept mastery of prospective physics teacher students who are designed based on the Bloom’s Taxonomy knowledge framework which includes four dimensions of knowledge (factual, conceptual, procedural, and metacognitive) that are spread in six dimensions of cognitive process dimensions (knowing, understanding, applying, analysing, evaluating and creating). The distribution of the EM-CMT items is shown in Table 1.

Pilot testing of the instrument test was imposed on prospective physics teacher students in one of the Universities in Makassar as many as 63 persons who had just completed a basic physics course. To achieve the research objectives, data analysis was carried out using the Rasch model application software program Winsteps version 3. 68.2 developed by Linacre [58] with dichotomous type. The selection of the Rasch dichotomy model is due to data from the EM-CMT instrument in the form of dichotomous data (multiple choice form), where there are only two possible scores obtained by respondents, namely a score of 0 and 1.

There are three criteria that need to be observed in determining the reliability of test instruments through the Rasch Model analysis, namely the Alpha Cronbach reliability value, the value of person and item reliability, and person and item separation [59] [60]. Determination of the test instrument validity with the Rasch Model analysis can be observed on three criteria, namely item measure, item fit, and item bias detection. The item measure or item difficulty level is obtained based on the results of the Winstep output analysis. Logit digit values generated from the Rasch Model analysis on measure items can provide indicators of the ability of respondents to answer items based on the level of difficulty of the items [57]. Rasch Model Analysis related to item fit level of was obtained based on the analysis results of misfit order statistical items. Item fit provides information to researchers whether the item functions normally to take measurements or not [59]. The criteria for determining item fit are based on three things, namely the value of outfit means-square (MNSQ), z-standard outfit value (ZSTD), and the value of point measure correlation (PT-MEASURE CORR). If one of these three criteria is not met, then it can be ascertained that the item is not good so it needs to be revised [38] [40] [41]. Three criteria for determining item fit have a range of digit values that must be met, namely: 1) the value of the mean square outfit (mnsq) received: 0.5 <mnsq <1.5; 2) received Z-standard (Zstd) outfit values: -2.0 <zstd <+2.0; and 3) point measure correlation value (Pt mean corr): 0.4 <Pt measure corr <0.85 [40] [41] [61].

Bond & Fox [37] [38] [39] asserted that the value of outfit means-square (MNSQ) can provide information to researchers about the suitability of items in the measurement of validity, while the value of point measure correlation (PT-MEASURE CORR) informs the extent to which the construct of the constructions is received in accordance with the objectives of its development. A positive point measure correlation (PT-MEASURE CORR) indicates that the item measures the construct to be measured, whereas a negative point measure correlation (PT-MEASURE CORR) indicates nonconformity. On the other hand, ZSTD is a hypothesis t test that can inform researchers whether the data fit perfectly to the model.

Other information that can be obtained from Rasch analysis is person fit. Boone [40] revealed that Rasch modeling can identify a person fit based on different response patterns. A different response pattern is the mismatch of answers given based on its ability compared to the ideal model. The criteria used to check person misfit are the same as the criteria used on item fit, which is based on a comparison of measurement results with the outfit means-square (MNSQ) value, z-standard outfit value (ZSTD), and point measure correlation value (PT-MEASURE) CORR) [41] [62]. If one of these three criteria is not met, then it can be ascertained that the item is not good so it needs to be revised [38] [40] [41]. The three criteria for determining person fit have a range of digit values that must be met, namely: 1) the value of the mean square outfit (mnsq) received: 0.5 <mnsq <1.5; 2) received Z-standard (Zstd) outfit values: -2.0 <zstd <+2.0; and 3) point measure correlation value (Pt mean corr): 0.4 <Pt measure corr <0.85 [40] [41] 61] [62]. A high Zstd outfit score (> 2.0) compared to a high MEASURE score can indicate that students with high ability answer incorrectly on easy items. Conversely, a high Zstd outfit value (> 2.0) compared to a low MEASURE score indicates that students with low ability answer correctly on difficult items but answer incorrectly on some items with moderate to lower levels of difficulty [63].

Other information that can be used from Rasch modeling in analyzing the validity of instrument items is a variable map or other terms Wright map or Person-Item map. Variable mapdescribes the distribution of respondents' abilities and the level of item difficulty with the same scale [38]. In the variable map, the respondent's ability is to the left of the map where at the bottom shows individuals with low abilities and the upper side of the map shows the position of students with high abilities. Meanwhile, the right part of the map illustrates the difficulty of the items. The lower part of the map shows the low difficulty items and the upper one for difficult items. In other words, the higher the logit value indicates the higher the respondent's ability and the more difficult the items and vice versa [41].

Table 1. Reliability and Categorize in Rasch Analysis

Statistics	Interval value of Coefficient	Category
Cronbach's Alpha	0,8 ≤ α	Very High
	0,7 ≤ α < 0,8	High
	0,6 ≤ α < 0,7	Good
	0,5 ≤ α < 0,6	Moderate
	α < 0,5	Low
Item and Person Reliability	0,94 ≤ r	Excellent
	0,91 ≤ r < 0,94	Very Good
	0,80 ≤ r < 0,91	Good
	0,67 ≤ r < 0,80	Sufficient
	r < 0,67	Low
Item and Person Separation		High separation value indicates that the instrument has a good quality since it can identify the group of item and person (respondent)

Another observation aspect that can be used in the Rasch analysis is an instrument's unidimensionality. The purpose of evaluating unidimensionality of an instrument is to ascertain whether the instrument measures what should be measured [42] [59], in this case, the construct of the electricity and magnetism concept mastery. Rasch analysis uses Principal Component Analysis (PCA) of standardized residues to measure the extent to which the diversity of instruments measures what is to be measured. The unidimensionality criteria based on the raw variance explained by the steps of the standard residual variant [59]. The raw variant values are explained sequentially that higher than 20% is acceptable, higher than 40% good, while higher than 60% is very good. Meanwhile, the ideal value for an unexplained 'variance must not exceed 15%.

RESULTS AND DISCUSSIONS

The results and discussion of this study were divided into five sections based on the characteristics of the Rasch analysis which were assessed in observing the reliability and validity of items in the EM-CMT instrument. Each of the characteristics of the Rasch analysis will be described below.

Blue Print of EM-CMT instrument
EM-CMT instrument was developed based on nine sub topics related to electrical and magnetic material, namely electric current, resistance, resistivity, Ohm's law, electric motion voltage, energy and electrical current conductors, resistors in series, parallel, and mixtures, Kirchhoff's law, RC charging and discharging capacitor circuits, magnetic fields, magnetic forces, and magnetic field sources. The development of this test instrument refers to the learning outcome rather than the semester lecture plan (RPS) that has been based on the cognitive process dimension framework of the Taxonomy Bloom model [64]. Meanwhile, the subject matter was developed by adapting various reference sources related to electricity and magnetism at the university level adapted from various sources on electricity and magnetism [65] [66] [67] [68]. Test instruments developed in the form of multiple choice as many as 40 items with the number of answer choices as much as five and the duration of the work time is 90 minutes. The distribution of material and item number is shown in Table 2.

Table 2. Blue Print of EM-CMT Instrument

Material	Topic	Number of Item
Current and Resistance	Electric Current	1, 2
	Resistance	3, 4
	Resistivity	5, 6
	Ohm’s Law	7
	Electric Voltage	8, 9
	Energy and Electric Power	10, 11, 12
Direct Current Circuit	Resistor in series, parallel and mixed circuits	13, 14, 15
	Kirchhoff’s Law	16, 17, 18, 19
	RC charging and discharging capacitor circuit	20, 21, 22, 23
Magnetism	Magnetic field	24, 25, 26, 27, 28
	Magnetic Force	29, 30, 31, 32, 33, 34, 39
	Sources magnetic field	35, 36, 37, 38, 39, 40

All items of the EM-CMT instrument developed were distributed in the dimensions of factual, conceptual, and procedural knowledge with levels of cognitive processes from the level of understanding (C2) to evaluating (C5).

The Rasch Analysis of Reliability and Item and Person Separation of the EM-CMT Instrument
Analysis on the level of reliability, separation of items and person of the EM-CMT instrument was obtained from the output of data processing using Winsteps program 3.68.2 version. There are three reviews of the reliability of the test instrument in the Rasch analysis, namely the Alpha Cronbach reliability value (KR-20), the value of person reliability, and the value of item reliability. For the observation of separation variables, item and person can be observed. The results of the analysis of several aspects of the observation of reliability and separation are shown in Table 3.

Table 3. The Value for Person and Item Reliability and Person and Item Separation

Statistic	Statistic Aspect	Value
Reliability	Alpha Cronbach	0.70
	Person reliability	0.69
	Item reliability	0.93
Separation	Person separation	1.50
	Item separation	3.58

Table 3 shows that the Cronbach Alpha reliability value (KR-20) is 0.70. With reference to Table 1 on the categorization of reliability and separation, the Alpha Cronbach (KR-20) reliability value indicates that EM-CMT instrument has good internal consistency reliability [69]. Meanwhile, Bond & Fox [38] argued that the Cronbach Alpha coefficient value obtained through the Rasch analysis approach in the range of 0.70 to 0.99 is the acceptable value with the best category. Person reliability obtained from the Rasch analysis is 0.69 and a person separation is 1.50. This person reliability value indicates the quite good category [69]. The value of the reliability coefficient obtained indicates that the response from respondents is quite good and consistent [39]. For the aspect of person separation, the coefficient value obtained is 1.50. The value of person separation must be greater than 1.00 to ensure that the respondents measured are spread as a whole [70]. Person separations of 1.50 (<3.0) fall into the acceptable category although this value indicates the test instrument is less sensitive to distinguish between high-ability and low-ability people [41]. The other important information obtained from the Rasch analysis is the value of item reliability of 0.93 and item separation of 3.58 (> 3.0). The value of this reliability item shows a very good category [69]. The coefficient value of the item separation obtained is in the good category according to Linacre [58] which confirms that the item separation value greater than 2.00 is interpreted as good. This implies that the sample person is sufficient to confirm the hierarchy of item difficulties [41] [70].

Item Fit Analysis of the EM-CMT

The criteria for determining item fit are based on three aspects, namely the value of Outfit means-square (MNSQ), Outfit z-standard value (ZSTD), and the value of point measure correlation (PT-MEASURE CORR). If one of these three criteria is not met, then it can be ascertained that the item is not good so it needs to be revised [38] [40] [41]. Table 4 shows a summary of the statistics of the fit items of the EM-CMT instrument. If the item does not meet one of these three criteria (outfit mnsq, outfit z-std, and Pt measure corr.) Then the item needs to be revised or discarded.

Table 4. Statistic Summary of the EM-CMT Instrument Item Fit.

No.	Item	Infit		Outfit		Pt-Measure	
		Mnzq	Zstd	Mnzq	Zstd	Corr.	Exp.
1.	S1	0,93	-0,2	0,73	-0,7	0,30	0,16
2.	S2	1,04	0,5	1,04	0,4	0,25	0,31
3.	S3	0,95	0,0	0,66	-0,5	0,24	0,11
4.	S4	0,87	-2,0	0,82	-1,7	0,47	0,28
5.	S5	1,08	1,1	1,09	0,8	0,16	0,28
6.	S6	0,80	-3,2	0,76	-2,4	0,57	0,28
7.	S7	1,00	0,2	0,94	0,1	0,11	0,10
8.	S8	1,03	0,3	0,99	0,00	0,29	0,32
9.	S9	0,93	-0,4	0,93	-0,4	0,41	0,32
10.	S10	0,99	0,0	1,00	0,0	0.32	0,31

No.	Item	Infit		Outfit		Pt-Measure	
		Mnzq	Zstd	Mnzq	Zstd	Corr.	Exp.
11.	S11	1,03	0,3	1,09	0,51	0,26	0,32
12.	S12	1,20	2,8	1,24	2,4	0,00	0,29
13.	S13	0,94	0,0	0,61	-0,6	0,27	0,11
14.	S14	1,01	0,1	0,94	-0,2	0,21	0,21
15.	S15	0,96	-0,5	0,94	-0,6	1,37	0,31
16.	S16	0,89	-0,9	0,91	-0,6	0,45	0,32
17.	S17	1,09	0,7	1,42	1,9	0,01	0,22
18.	S18	1,01	0,2	1,02	0,2	0,26	0,28
19.	S19	1,01	0,1	1,01	0,2	0,27	0,29
20.	S20	1,10	1,5	1,37	2,9	0,07	0,28
21.	S21	0,96	-0,5	0,94	-0,5	0,37	0,31
22.	S22	1,12	0,6	1,33	1,3	0,10	0,31
23.	S23	0,99	-0,1	0,98	-0,1	0,33	0,31
24.	S24	0,91	-0,3	0,78	-0,7	0.33	0,18
25.	S25	0,89	-0,7	0,89	-0,6	0,47	0,32
26.	S26	0,97	-0,4	0,95	-0,4	0,33	0,28
27.	S27	0,76	-1,0	0,66	-1,2	0,63	0,30
28.	S28	1,10	1,0	1,19	1,7	0,15	0,31
29.	S29	1,01	0,1	1,11	0,6	0,27	0,32
30.	S30	1,00	0,0	1,00	0,1	0,28	0,28
31.	S31	1,02	0,2	1,00	0,00	0.30	0,32
32.	S32	1,02	0,3	1,02	0,2	0.28	0,31
33.	S33	0,88	-0,5	0,93	-0,2	0,43	0,31
34.	S34	0,99	0,0	0,99	-0,1	0,31	0,31
35.	S35	1,05	0,7	1,00	0,0	0,23	0,28
36.	S36	1,05	0,5	1,07	0,6	0,24	0,31
37.	S37	1,03	0,3	1,05	0,3	0,27	0,32
38.	S38	1,11	1,0	1,11	0,8	0,17	0,32
39.	S39	1,03	0,2	1,03	0,3	0,28	0,32
40.	S40	0,99	0,0	0,98	-0,1	0,32	0,31

Table 4 shows that S12 and S20 items have a tendency to not fit because they do not meet the requirements in Outfit Zstd (for S12 items the value is 2.4 and S20 has a value of 2.9) and Pt. measure corr. value (S12 is 0.00 and S20 is 0.07), but for Outfit Mnsq criteria both items are still within the allowed limits so item S12 and S20 are retained. For item S6 it just doesn't meet Outfit Zstd requirements (value -2.4) so it doesn't need to be changed. As with a number of other items that only do not meet one of the three criteria, namely Pt. measure corr. (all items except items S4, S9, S16, S25, S27, and S33). However, the items referred to fulfill the criteria for Outfit Mnzq and Outfit Zstd so that it does not need to be changed. Meanwhile, six other items (S4, S9, S16, S25, S27, and S33) have fulfilled all three criteria, so there are no problems with the six items. Thus, the final conclusion there are no questions that need to be changed or replaced.

Person Fit Analysis of the EM-CMT

Other information that can be used to observe items that do not fit the model (misfit), namely: 1) outfit mean square value (mnsq) received: 0.5 <mnsq <1.5; 2) received outfit Z-standard (Zstd) values: -2.0 <zstd <+2.0; and 3) point measure correlation value (Pt mean corr): 0.4 <Pt measure corr <0.85 [40] [41] [61]. Using the three criteria for person fit observation, there are several respondents who experienced misfits and are summarized in Table 5.

Table 5. Misfit Orders for Respondents (Persons) in the EM-CMT Instrument

No.	Person	Score Total	MEASURE	Outfit MNSQ (0,5 – 1,5)	Outfit ZSTD (-2,0 – 2,0)	PT-Measure Corr. (0,4 – 0,85)
1.	R62	13	-0,84	2,07	3,2	0,02
2.	R51	14	-0,69	1,72	2,5	0,09
3.	R22	21	0,22	1,70	2,4	0,26
4.	R36	14	-0,69	1,62	2,2	0,11
5.	R34	13	-0,84	1,58	1,9	0,19
6.	R61	16	-0,42	1,45	1,8	0,28
7.	R08	31	1,56	1,28	0,7	0,17
8.	R02	29	1,25	1,10	0,4	0,20
9.	R33	15	-0,55	1,20	0,9	0,36
10.	R25	18	-0,16	1,17	0,8	0,38
11.	R59	14	-0,69	1,14	0,6	0,39
12.	R52	14	-0,69	1,14	0,6	0,38
13.	R05	24	0,59	1,07	0,3	0,30
14.	R09	36	2,61	1,10	0,4	0,13
15.	R19	24	0,59	0,97	0,0	0,35
16.	R01	31	-0,16	0,93	0,0	0,27
17.	R30	30	1,40	1,03	0,2	0,29

Table 5 shows the respondent (R) whose response to the item experienced misfit based on Rasch analysis. In other words, the respondent (person) gives an answer that is not suitable with his ability compared to the ideal model. All persons (responses) are given the initial code "R". Based on Table 5, there are four respondents (R62, R51, R22, R36) who do not meet the three criteria for determining the suitability of an incompatible item (misfit), that is outside the MNSQ outfit, ZSTD outfit, and PT-Measure Corr. R34 respondents only fulfill ZSTD outfit aspects (+1.9) while MNSQ and PT-Measure Corr outfit aspects are outside the allowed limits. Some respondents who met the requirements in the MNSQ outfit and ZSTD outfit aspects but the PT-Measure Corr value was outside the allowed limits, namely R61, R08, R02, R33, R25, R59, R52, R05, R09, R19, R01, and R30. Meanwhile, other respondents have answer patterns with all three criteria meeting the requirements. Based on the analysis of misfit order person fit on outfit MNSQ, outfit ZSTD, and PT-Measure Corr criteria, it can be concluded that there were seventeen out of sixty-three respondents (people) who experienced unusual response patterns.

Variable Map (Wright Map/Person-Item Map)

Figure 1 shows a variable map (Wright map) that shows the distribution of people and items in the logit measurement scale. Map variables provide meaningful information on how the distribution of the difficulty level of items corresponds to the level of ability of the person [69].

The Wright map in Figure 1 provides information that R09 is the highest skilled respondent. It can be seen that respondents R09 all questions can be done correctly because their logit position is at a higher level than all items. The lowest ability respondents are R12, R34, R62, and R63 at the same level. However, these four respondents (R12, R34, R62, and R63) have the ability to correctly answer a number of items (S1, S3, S7, S14, S17, and S24 items). This is due to the logit position of the ability of the four respondents is higher than the logit position of these items.

The logit value of two respondents namely R01 and R08 occupies the same logit position and has a lower ability than R09 respondents as the respondents with the highest ability. It can be seen that the logit position of the two respondents compared to logit item S22, almost has the same logit value position. This condition explains that both respondents had a 50% chance of working on S22 questions correctly. However, these two respondents did not have the ability to answer correctly for items S27 and S33 because the position of the logit value of the two respondents was lower than the logit value position of items S27 and S33. The condition of the distribution of respondents' abilities and the ability of items shows that there is a proportion so that it can be concluded that all items do not need to be discarded / replaced.

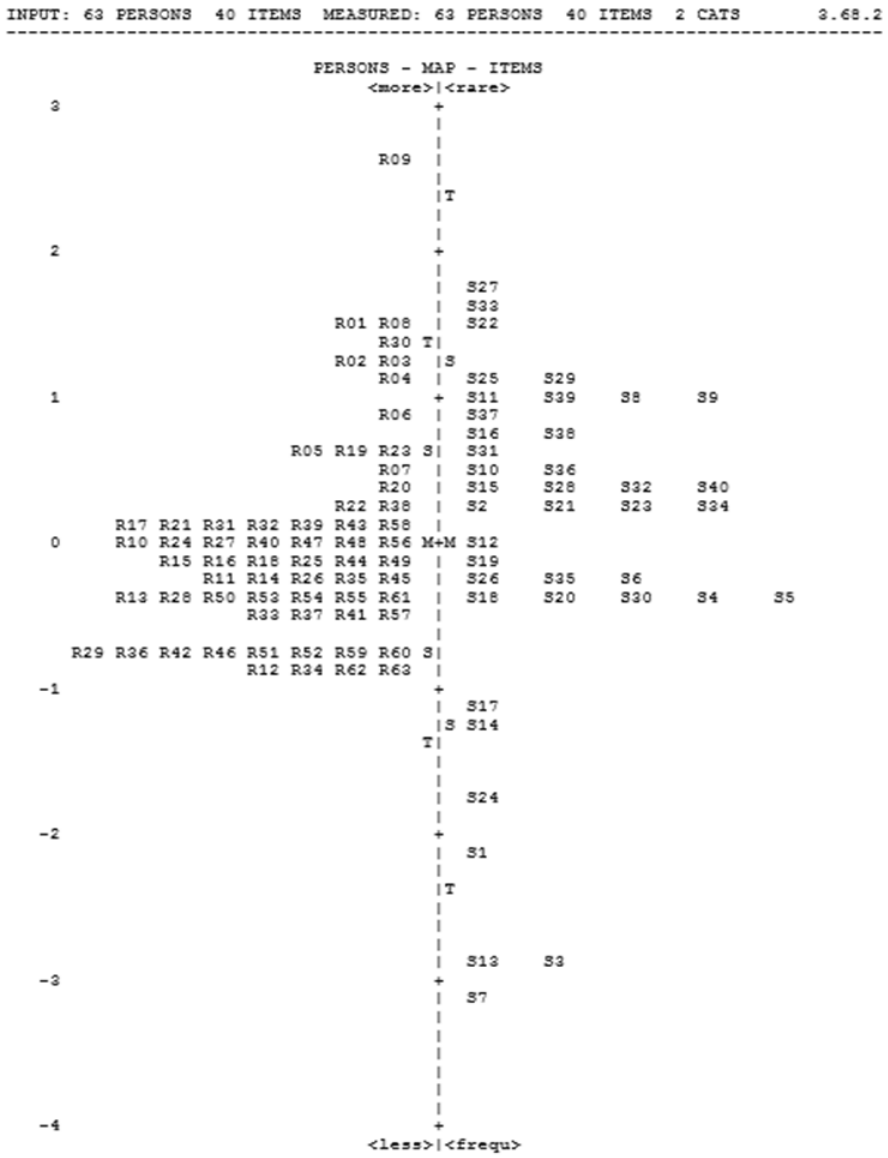

Fig 1. *Variable Map of Person and Item*

Unidimensionality Analysis of the EM-CMT

Observation of unidimensionality aspect is done based on the criteria for the percentage value of variance.Figure 2 shows the output of the Rasch analysis results on the unidimensioanlity aspect.

```
INPUT: 63 PERSONS  40 ITEMS  MEASURED: 63 PERSONS  40 ITEMS  2 CATS       3.68.2
--------------------------------------------------------------------------------

     Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                                              -- Empirical --    Modeled
Total raw variance in observations     =       53.5 100.0%        100.0%
  Raw variance explained by measures   =       13.5  25.3%         21.7%
    Raw variance explained by persons  =        3.2   6.0%          5.1%
    Raw Variance explained by items    =       10.3  19.3%         16.6%
  Raw unexplained variance (total)     =       40.0  74.7% 100.0%   78.3%
    Unexplned variance in 1st contrast =        3.7   6.9%   9.2%
    Unexplned variance in 2nd contrast =        3.1   5.8%   7.8%
    Unexplned variance in 3rd contrast =        2.8   5.3%   7.1%
    Unexplned variance in 4th contrast =        2.5   4.7%   6.2%
    Unexplned variance in 5th contrast =        2.1   4.0%   5.3%
```

Fig 2. *Standardized Residual Variance*

Figure 2 shows percentage value of "raw variance explained by measure" is 25.3%. Based on the quality categorization of the percentage aspects of raw variance explained by measure determined by Sumintono & Widhiarso [69], the value of "raw variance explained by measure" is in the acceptable category. In addition, the percentage value of the unexplained variance from the first contrast to the fifth shows less than 10% which falls in the range less than the ideal range value of less than 15%.

CONCLUSION AND SUGGESTION

The results of the reliability and validity analysis of the Rasch analysis show that the concept mastery test instrument developed has fulfilled the requirements in all aspects of observation, namely (1) the value of the Alpha Cronbach reliability level with good and acceptable categories (0.70); (2) the value of the reliability of the person is in the quite good category (0.69); (3) the personnel separation coefficient of 1.50 which indicates the response of the respondents is quite good and consistent; (4) analysis of observational aspects of item fit shows that no items need to be changed or removed; (5) for the aspect of person fit, the results of the analysis showed that there were seventeen out of sixty three respondents who experienced unusual response patterns; (6) a review of aspects of observing the variable map showing the distribution of respondents' abilities and items already proportions; (7) analysis of unidimensional aspect shows the value of "raw variance explained by measure" is in the acceptable category. Based on the results of the analysis on a number of aspects, it can be concluded that the instrument is an appropriate test instrument to use.

ACKNOWLEDGMENTS

This study is supported by Department of Physics Education, Faculty of Teacher Training and Education, Muhammadiyah University of Makassar. We thank to the research institute LP3M Muhammadiyah University of Makassar which has provided research funding assistance in the research collaboration between junior high school science teacher and physics education lecturer scheme. We also are grateful to the participants who have been contributed in this study.

DOI: 10.26737/jipf.v8i2.3877

REFERENCES

[1] Shipstone, D. M. (1984). A study of children's understanding of electricity in simple DC circuits. *European journal of science education*, *6*(2): 185-198.

[2] Shipstone, D. (1988). Pupils' understanding of simple electrical circuits. Some implications for instruction. *Physics education*, *23*(2): 92.

[3] Arnold, M., & Millar, R. (1987). Being constructive: An alternative approach to the teaching of introductory ideas in electricity. *International Journal of Science Education*, *9*(5): 553-563.

[4] Osborne, R. J., & Cosgrove, M. M. (1983). Children's conceptions of the changes of state of water. *Journal of research in Science Teaching*, *20*(9): 825-838.

[5] Osborne, R. (1983). Towards modifying children's ideas about electric current. *Research in Science & Technological Education*, *1*(1): 73-82.

[6] Hekkenberg, A., Lemmer, M., & Dekkers, P. (2015). An analysis of teachers' concept confusion concerning electric and magnetic fields. *African Journal of Research in Mathematics, Science and Technology Education*, *19*(1): 34-44.

[7] Tarciso Borges, A., & Gilbert, J. K. (1999). Mental models of electricity. *International journal of science education*, *21*(1): 95-117.

[8] Cosgrove, M. (1995). A study of science-in-the-making as students generate an analogy for electricity. *International journal of science education*, *17*(3): 295-310.

[9] Cohen, R., Eylon, B., & Ganiel, U. (1983). Potential difference and current in simple electric circuits: A study of students' concepts. *American journal of Physics*, *51*(5): 407-412.

[10] Paatz, R., Ryder, J., Schwedes, H., & Scott, P. (2004). A case study analysing the process of analogy-based learning in a teaching unit about simple electric circuits. *International Journal of Science Education*, *26*(9): 1065-1081.

[11] Psillos, D., Koumaras, P., & Valassiades, O. (1987). Pupils' representations of electric current before, during and after instruction on DC circuits. *Research in Science & Technological Education*, *5*(2): 185-199.

[12] Engelhardt, P. V., & Beichner, R. J. (2004). Students' understanding of direct current resistive electrical circuits. *American journal of physics*, *72*(1): 98-115.

[13] Finkelstein, N. (2005). Learning physics in context: A study of student learning about electricity and magnetism. *International Journal of Science Education*, *27*(10): 1187-1209.

[14] Zacharia, Z. C., & De Jong, T. (2014). The effects on students' conceptual understanding of electric circuits of introducing virtual manipulatives within a physical manipulatives-oriented curriculum. *Cognition and instruction*, *32*(2): 101-158.

[15] Stocklmayer, S. M., & Treagust, D. F. (1996). Images of electricity: How do novices and experts model electric current?. *International Journal of Science Education*, *18*(2): 163-178.

[16] Heller, P. M., & Finley, F. N. (1992). Variable uses of alternative conceptions: A case study in current electricity. *Journal of Research in Science Teaching*, *29*(3): 259-275.

[17] Heywood, D., & Parker, J. (1997). Confronting the analogy: primary teachers exploring the usefulness of analogies in the teaching and learning of electricity. *International Journal of Science Education*, *19*(8): 869-885.

[18] Dega, B. G., Kriek, J., & Mogese, T. F. (2013). Students' conceptual change in electricity and magnetism using simulations: A comparison of cognitive perturbation and cognitive conflict. *Journal of Research in Science Teaching*, *50*(6): 677-698.

[19] 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.

[20] Groundlund, R. L., & Linn, N.E. (1990). *Measurement and Evaluation in Teaching*. Prentice Hall College Div.

[21] Thorndike, R. L. (1971). *Educational Measurement, Second Ed*. American Council on Education.

[22] Aydın, U., & Ubuz, B. (2015). The thinking-about-derivative test for undergraduate students: development and validation. *International Journal of Science and Mathematics Education*, *13*(6): 1279-1303.

[23] Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items*. Routledge.

[24] Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-

writing guidelines for classroom assessment. *Applied measurement in education*, *15*(3): 309-333.
[25] Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and Validating Test*. Routledge Taylor & Francis Group.
[26] Groundlund, N. E. (1982). *Constructing Achievement Test, Third Ed*. Prentice-Hall, Inc.
[27] Groundlund, N. E. (2003). *Assessment of Student Achievement, Seventh Ed*. Pearson Education, Inc.
[28] Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*. Longman Publishing.
[29] Russell, M. K., Airasian, P. W. (1967). *Classroom Assessment: Concepts and Applications, 7th ed*. McGraw-Hill Companies.
[30] Mehrens, W. A. & Lehmann, I. J. (1991). *Measurement and Evaluation in Education and Psychology, Fourth Ed*. Wadsworth/Thomson Learning.
[31] Skorupiński, P. M. (2015). American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Standards for educational and psychological testing. *Kwartalnik Pedagogiczny*, *238*(4): 201-203.
[32] Susongko, P. (2016). Validation of science achievement test with the rasch model. *Jurnal Pendidikan IPA Indonesia*, *5*(2): 268-277.
[33] Retnawati, H. (2016). *Validitas, Reliabilitas, & Karakteristik Butir (Panduan untuk Peneliti, Mahasiswa, dan Psikometrian)*. Parama Publishing.
[34] Hambleton, R. K. & Swaminatan, H. (1985). *Item Response Theory: Principles and Applications*. Springer Science+Business Media, LLC.
[35] Hambleton, R. K., Swaminatan, H., & Rogers, H. J. (1991). *Fundamentals of Item Respons Theory*. Sage Publication, Inc..
[36] Demars, C. (2010). *Item Respons Theory: Understanding Statistics Measurement*. Oxford University Press.
[37] Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Lawrence Erlbaum Associates.
[38] Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Science, Third Edit*. Routledge Taylor & Francis Group.
[39] Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edi*. Lawrence Erlbaum Associates.
[40] Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how?. *CBE—Life Sciences Education*, *15*(4): 1-7.
[41] Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer.
[42] Abdaziz, A., Jusoh, M. S., Amlus, H., & Salleh, T. S. (2014). Construct validity: A Rasch measurement model approaches. *Journal of Applied Science and Agriculture*, *9*(12): 7-12.
[43] Cheng, M. M., & Oon, P. T. (2016). Understanding metallic bonding: Structure, process and interaction by Rasch analysis. *International Journal of Science Education*, *38*(12): 1923-1944.
[44] Gänswein, W. (2011). *Effectiveness of Information Use for Strategic Decision Making*. Gabler.
[45] Luo, T., Wang, J., Liu, X., & Zhou, J. (2019). Development and application of a scale to measure students' STEM continuing motivation. *International Journal of Science Education*, *41*(14): 1885-1904.
[46] Tabatabaee-Yazdi, M., Motallebzadeh, K., Ashraf, H., & Baghaei, P. (2018). Development and Validation of a Teacher Success Questionnaire Using the Rasch Model. *International Journal of Instruction*, *11*(2): 129-144.
[47] Yang, W., & Liu, E. (2016). Development and validation of an instrument for evaluating inquiry-based tasks in science textbooks. *International Journal of Science Education*, *38*(18): 2688-2711.
[48] Commons, M. L., Goodheart, E. A., Pekker, A., Dawson, T. L., Draney, K., & Adams, K. M. (2008). Using Rasch scaled stage scores to validate orders of hierarchical complexity of balance beam task sequences. *Journal of Applied Measurement*, *9*(2): 182-199.
[49] Fulmer, G. W. (2015). Validating proposed learning progressions on force and motion using the force concept inventory: Findings from Singapore secondary schools. *International Journal of Science and Mathematics Education*, *13*(6): 1235-1254.

[50] bin Khairani, A. Z., & bin Abd Razak, N. (2012). Advance in educational measurement: A Rasch model analysis of Mathematics proficiency test. *International Journal of Social Science and Humanity*, *2*(3): 248.

[51] Kuo, C. Y., Wu, H. K., Jen, T. H., & Hsu, Y. S. (2015). Development and validation of a multimedia-based assessment of scientific inquiry abilities. *International Journal of Science Education*, *37*(14): 2326-2357.

[52] Van Zile-Tamsen, C. (2017). Using Rasch analysis to inform rating scale development. *Research in Higher Education*, *58*(8): 922-933.

[53] Briggs, A. L., Jackson, T. R., Bruce, S., & Shapiro, N. L. (2005). The development and performance validation of a tool to assess patient anticoagulation knowledge. *Research in social and administrative Pharmacy*, *1*(1): 40-59.

[54] Nielsen, J. B., Kyvsgaard, J. N., Sildorf, S. M., Kreiner, S., & Svensson, J. (2017). Item analysis using Rasch models confirms that the Danish versions of the DISABKIDS® chronic-generic and diabetes-specific modules are valid and reliable. *Health and Quality of Life Outcomes*, *15*(1): 1-10.

[55] Tran, V. D., Dorofeeva, V. V., & Loskutova, E. E. (2018). Development and validation of a scale to measure the quality of patient medication counseling using rasch model. *Pharmacy Practice (Granada)*, *16*(4).

[56] Runnels, J. (2012). Using the Rasch model to validate a multiple choice English achievement test. *International Journal of Language Studies*, *6*(4): 141-155.

[57] Saidi, S. S., & Siew, N. M. (2019). Reliability and Validity Analysis of Statistical Reasoning Test Survey Instrument Using the Rasch Measurement Model. *International Electronic Journal of Mathematics Education*, *14*(3): 535-546.

[58] Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of applied measurement*, *3*(1): 85-106.

[59] Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim komunikata.

[60] Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)*. Trim Komunikata Publishing House.

[61] Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, *90*(2): 253-269.

[62] Davidowitz, B., & Potgieter, M. (2016). Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry. *International Journal of Science Education*, *38*(9): 1483-1503.

[63] Nevin, E., Behan, A., Duffy, G., Farrell, S., Harding, R., Howard, R., ... & Bowe, B. (2015, July). Assessing the validity and reliability of dichotomous test results using Item Response Theory on a group of first year engineering students. In *The 6th Research in Engineering Education Symposium (REES 2015), Dublin, Ireland, July* (pp. 13-15).

[64] Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. David McKay Company, Inc.

[65] Giancoli, D. C. (2016). *Physics Principles With Applications, Global Ed*. Pearson.

[66] Serway, R. A., & Jewett, J. W. (2014). *Physics for Scientists and Engineers with Modern Physics, Ninth Ed*. Brooks/Cole Cengage Learning.

[67] Walker, J., Halliday, D., & Resnick, R. (2014). *Fundamental Physics, Tenth Ed*. John Wiley & Sons, Inc.

[68] Rahmawati, R., Rustaman, N. Y., Hamidah, I., & Rusdiana, D. (2018). The development and validation of conceptual knowledge test to evaluate conceptual knowledge of physics prospective teachers on electricity and magnetism topic. *Jurnal Pendidikan IPA Indonesia*, *7*(4): 283-490.

[69] Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch Pada Asesmen Pendidikan, I*. Penerbit Trim Komunikasi.

[70] Krishnan, S., & Idris, N. (2014). Investigating reliability and validity for the construct of inferential statistics. *International Journal of Learning, Teaching and Educational Research*, *4*(1): 51-60.