Phys. Comm. 7 (2) 2023 : 43 - 53

Physics Communication

http://journal.unnes.ac.id/nju/index.php/pc

Rasch Model Analysis: Development of Literacy Numeracy Assessment Instrument of Electromagnetic Induction

Nurul Apriani Susanti✉, Sarwi, Mahardika Prasetya Aji
Postgraduate Universitas Negeri Semarang, Semarang, Indonesia

Article Info

Article history:
Submitted
Revised
Published

Keywords:
Minimum Capability Assessment (MCA), Electromagnetic induction, Numeracy Literacy, Rasch Model Analysis, Winstep software

Abstract

This research aimed to describe the validity of a numeracy literacy assessment instrument of electromagnetic induction based on minimum capability assessment (MCA) which was develop using Rasch model analysis assisted by Winstep software. The research employed the research and development design with a modified Mardapi model. The instrument was constructed by developing fifteen questions. The analysis of the instrument was utilized to measure the construct and item validity, reliability, item difficulty level, and item discriminatory. This research involved 32 participants in the small-scale test and 206 participants in the large-scale trial. The test result showed that the fifteen questions were suitable to use because it has values that meet the criteria for all measurement variables both on small scale and large-scale tests and could be assess students' numeracy literacy skills.

Correspondence:
Postgraduate Universitas Negeri Semarang
Jalan Kelud Utara III No.37, Kota Semarang, Indonesia 50237
nurulapriani234@students.unnes.ac.id

p-ISSN 2528-5971
e-ISSN 2528-598X

INTRODUCTION

Assessment in the learning process is an important component of educational evaluation activities (Syadiah & Hamdu, 2020). Through the assessment, information about the competencies that have been achieved by students during the learning process will be obtained. The results of the assessment can also be an indicator of the quality of learning carried out by a teacher, so that in the future the teacher can make decisions regarding the implementation of further learning activity. The quality of the learning process can also be further improved to increase the competence possessed by students.

Numerical literacy is one of the basic competencies that must be possessed by students in the current industrial revolution era. Numerical literacy is not just the ability to count theoretically but also how to apply it in everyday life (Saputri & Diana, 2021). Numerical literacy is the knowledge and skills to use various numbers and symbols related to basic mathematics to solve practical problems in various contexts of daily life and analyze information presented in various forms (graphs, tables, charts, etc.) then use the interpretation of the analysis results to predict and make decisions in everyday life (Han, 2017; OECD, 2021). This can help students to recognize the role of mathematics in everyday life so that they can make the necessary judgments and decisions and responsible and think logically (Pusmenjar, 2020).

Numerical literacy is one of the abilities measured in the Minimum Competency Assessment (MCA). Minimum Competency Assessment, which is translated to Asesmen Ketuntasan Minimum in Bahasa Indonesia, is one of the three assessments contained in the National Assessment. MCA is intended to produce information about the competency level of students. These results can be used by teachers to develop effective and quality learning strategies according to the level of achievement of students. Thus, "*teaching at the right level*" can be applied.

The questions contained in the MCA are included in the type of *High Order Thinking Skill* (HOTS) questions. Questions of this type are still rarely used in learning assessment, especially those related to numeracy literacy skills. Even if it is applied in a sustainable manner, the ability of students can be further improved. To familiarize students with MCA model questions, teachers need to familiarize students with similar types of questions. However, the availability of MCA questions is still not varied, especially for physics lessons. The number is still very limited and only on certain materials, so it needs to be developed.

Assessment activities required measurement or test instruments that are expected to produce valid and reliable data (Mardapi, 2015). Therefore, it is necessary to analyze the questions that have been developed to determine their quality. The quality of an instrument can be seen from the validity and reliability values obtained from the measurement results (Himelfarb, 2019). Other variables that indicate the quality of the assessment instrument are the difficulty level of the items and discriminating power. The success rate of the questions in measuring the actual ability of students can be known through an analysis of the level of difficulty and discriminating power (Dewi et al., 2019). A good quality instrument is able to provide accurate information regarding students' abilities on the competency being tested (Azizah & Wahyuningsih, 2020).

The Rasch model is an analysis that can helps teachers, lecturers and educational assessment researchers to improve the quality of the analysis carried out, since it is not only used to analyze the quality of an instrument, but also able to analyze students' abilities (Aghekyan, 2020; Planinic et al., 2019; Suminto & Widhiarso, 2014). The Rasch model has the ability to measure the test questions that the teacher will give to their students by generating valid data and causing the question analysis or research activities to produce what is desired.

The advantages of developing questions using the Rasch Model will produce questions that are tested for their level of difficulty. The difficulty level of a question is the opportunity to answer a question correctly at a certain level of ability which is usually expressed in the form of an index. The function of item difficulty level is to associated with the purpose of the test as usual (Rangkuti, 2011). Learning assessments provide good information for teachers to help students to learn better.

Another advantage of Rasch analysis is to predict missing data, which is based on systematic response patterns. So, with the Rasch modelling analysis, it will produce a suitability statistical analysis that provides information to the user whether the data obtained ideally depicts that people who have high abilities provide patterns of answers to items according to their level of difficulty (Sumintono & Widhiarso, 2015).

Therefore, the aims of this research to describe the quality of the numeracy literacy assessment instrument of electromagnetic induction was developed based on MCA for physics lessons using the Rasch model analysis.

METHODS

This research used research and development design compiled with modified Mardapi (2017) model. There are nine stages carried out in this research. The development flow chart of this

research can be seen in Figure 1.

In this research, the developed product was an evaluation instrument of numeracy literacy. The instrument consisted of fifteen questions with numeracy literacy indicators. There were three indicators numeracy literacy, namely (1) Formulate a mathematical situation; (2) The process employs concepts, facts, procedures, and mathematical reasoning; and (3) The process of interpreting, applying, and evaluating mathematical results. Each indicator consists of five questions. Each question has indicators of achieving different competencies which are listed in Table 1.

Table 1. The Test Grid

No.	Ability aspect of Literacy Numeracy	Indicators of Competence Achievement	Quention Indicators	Cognitive Level
1.	Formulate a mathematical situation	Applying Faraday's law to analyze the Induction Electromotive Force (EMF) in some electromagnetic phenomena	Provided a mathematical formulation to determine the magnetic flux value, students are able to analyze the change of magnetic flux value.	C4
2.	Formulate a mathematical situation	Applying Faraday's law to analyze the Induction Electromotive Force (EMF) in some electromagnetic phenomena	Provided data from the results of the Faraday experiment, students are able to analyze the changes of EMF induction.	C4
3.	Formulate a mathematical situation	Identify the phenomenon of electromagnetic induction in everyday life.	Provided data on armature generator specifications, students can analyze the resulting EMF induction.	C4
4.	Formulate a mathematical situation	Applying Faraday's law to analyze the Induction Electromotive Force (EMF) in some electromagnetic phenomena	Provided a graph of the relationship between changes in magnetic field flux per unit time to the EMF induction of the experimental results, students can conclude the relationship between that.	C5
5.	Formulate a mathematical situation	Identify the phenomenon of electromagnetic induction in everyday life.	Provided reading material about bicycle dynamos, students can formulate the relationship between the amount of magnetic induction and the number of turns on the coil with the result of EMF induction	C6
6.	The process employs concepts, facts, procedures, and mathematical reasoning	Applying Faraday's law to analyze the Induction Electromotive Force (EMF) in some electromagnetic phenomena	Presented an image of Faraday's law experiments, students can analyze the direction and the result value of EMF induction current	C4
7.	The process employs concepts, facts, procedures, and mathematical reasoning	Applying Faraday's law to analyze the Induction Electromotive Force (EMF) in some electromagnetic phenomena	Provided with three simple electric circuits covered by a magnetic field, students can compare the result value of EMF induction	C5
8.	The process employs concepts, facts, procedures, and mathematical reasoning	Applying Faraday's law to analyze the Induction Electromotive Force (EMF) in some electromagnetic phenomena	Provided data on changes in the value of physical quantities on electric generators, students can compare the resulting maximum EMF induction values.	C5
9.	The process employs concepts, facts, procedures,	Identify the phenomenon of electromagnetic induction in everyday life.	Provided data specifications of a transformer, students can	C5

No.	Ability aspect of Literacy Numeracy	Indicators of Competence Achievement	Quention Indicators	Cognitive Level
	and mathematical reasoning		evaluate the work of the transformer.	
10.	The process employs concepts, facts, procedures, and mathematical reasoning	Identify the phenomenon of electromagnetic induction in everyday life.	Provided reading about the specifications of the transformer, students can reconstruct the type of transformer by changing the value of the quantity that affects it.	C4
11.	The process of interpreting, applying, and evaluating mathematical results	Identify the phenomenon of electromagnetic induction in everyday life.	Provided a picture of a transformer that connected to a simple electrical circuit, students can analyze the state of the circuit after changes are made.	C4
12.	The process of interpreting, applying, and evaluating mathematical results	Applying Faraday's law to analyze the Induction Electromotive Force (EMF) in some electromagnetic phenomena	Provided with pictures of the Faraday's law experimental, students can analyze the EMF induction and the resulting induced currents.	C4
13.	The process of interpreting, applying, and evaluating mathematical results	Applying Faraday's law to analyze the Induction Electromotive Force (EMF) in some electromagnetic phenomena	Several statements related to EMF induction value produced by a coil are presented, students can evaluate the correct statement.	C5
14.	The process of interpreting, applying, and evaluating mathematical results	Applying Faraday's law to analyze the Induction Electromotive Force (EMF) in some electromagnetic phenomena	Presented an image of a modified Faraday series of experiments, students can predict the result of EMF induction value.	C5
15.	The process of interpreting, applying, and evaluating mathematical results	Applying Faraday's law to analyze the Induction Electromotive Force (EMF) in some electromagnetic phenomena	Presented an image of an electric circuit in a magnetic field, students can predict the magnitude and direction of the induced current passing through the resistance.	C5

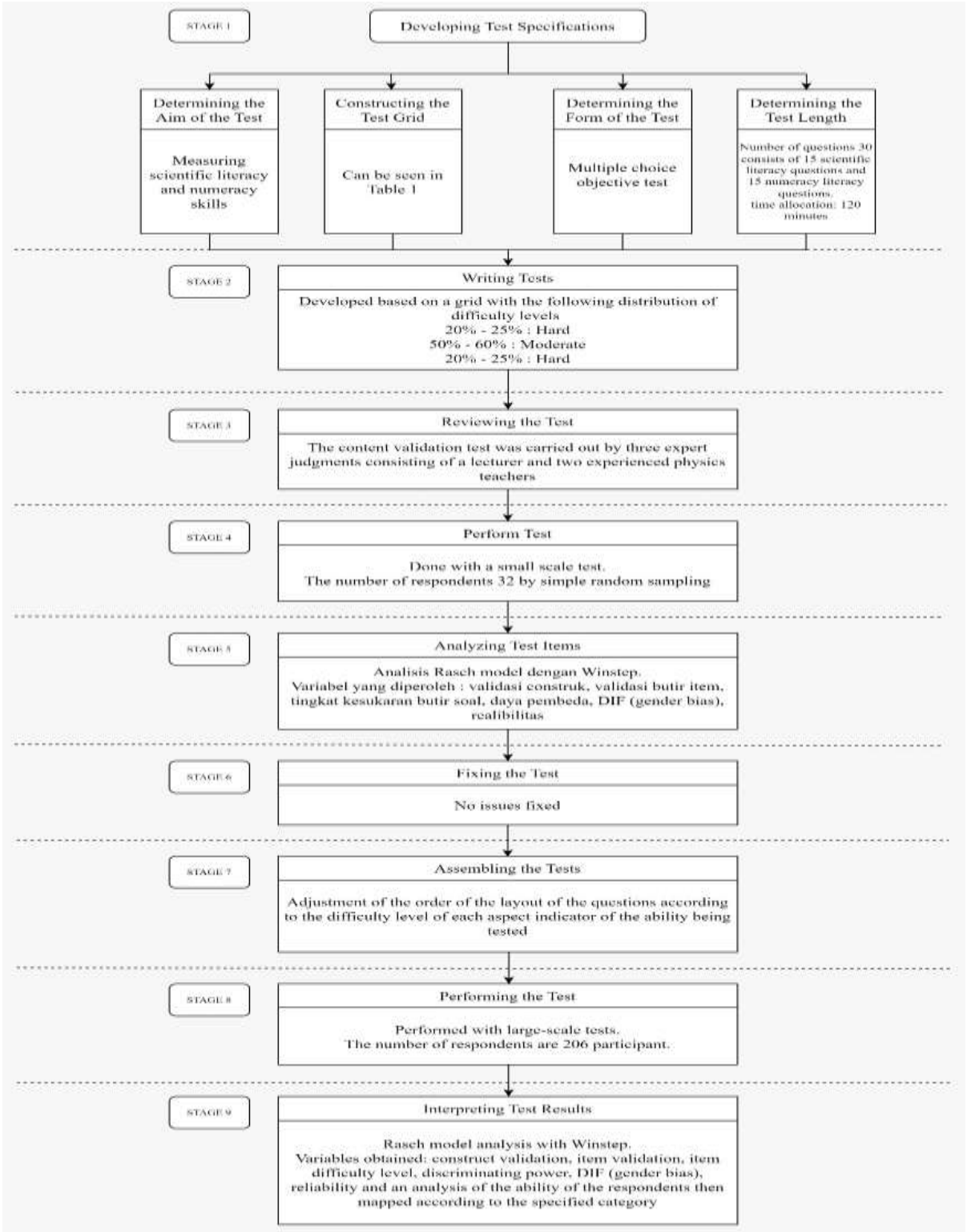


Figure1. Numeracy literacy assessment instrument of electromagnetic induction based on AKM developed flowchart

RESULTS AND DISCUSSION

3.1 Construct Validity

Construct validity refers to the ability of the test items to measure every aspect of thinking as stated in the specific instructional objectives (Arikunto, 2009). To get construct validity using the winstep application, it can be seen in the ***output table*** on main menu then selecting ***table 23. Item: dimensionality*****.** Validity is also known as

unidimensionality (Sumintono and Widhiarso, 2014: 122). Validity test based on item dimensionality by looking at raw variance explained by measures. The unidimensionality requirements are shown in Table 2. *Unexplained variance in constrast 1-5 interpreting* was showed in Table 3.

Table 2. Unidimentionality Interpreting

Interpreting	Score (%)
Qualify	20.1 – 39.9
Good	40 – 59.9
Excellent	60 – 100

Table 3. *Unexplained variance in constrast 1-5* interpreting

Interpreting	score (%)
Poor	> 15
Fair	10-15
Good	5-10
Very Good	3-5
Excellent	< 3

Table 4. Unidimentionality Result

Indicators	Small-scale		Large-scale	
	score (%)	Interpreted	score (%)	Interpreted
Raw variance explained by measure	31.5	Qualify	29.1	Qualify
unexplained variance in 1st contrast	11.8	Fair	9.6	Good
unexplained variance in 2nd contrast	10.1	Fair	8.2	Good
unexplained variance in 3rd contrast	8.3	Good	7.1	Good
unexplained variance in 4th contrast	7.5	Good	6.6	Good
unexplained variance in 5th contrast	6.0	Good	5.4	Good

The raw variance in observations obtained in the small-scale and large-scale trials were not much different, namely 31.5% and 29.1%. This shows that the AKM-based numeracy literacy assessment instrument that was developed is included in the "**Qualify**" category based on Table 1 so that it can be said that the item items are valid. The low value of raw variance in observations, for both small-scale tests and large-scale tests, where the value is less than 40% (<40%) indicates a not-so-good distribution of item difficulty levels. Nonetheless, it can be said that the instruments developed have been able to measure students' numeracy literacy skills in accordance with the basic competencies and competency indicators developed on electromagnetic induction material.

Other data that can be explained is the unexplained variance value which ideally has a value not exceeding 15% for each variant and the eigenvalue on the unexplained variance in 1st contrast is not more than 3. Based on Table 4 on a small-scale test, the value of each unexplained variance meets the criteria as shown in Table 3, which is between 6.0% - 11.8% and the eigenvalue for unexplained variance in 1st contrast is 2.6. Likewise with the value of each unexplained variance obtained in a large-scale test with a value range of 5.4% - 9.6% and the eigenvalue for unexplained variance in 1st contrast is 2.0. This shows that the level of independence of the items in the instrument is good, the items are not affected by other dimensions. The construct validity analysis carried out by researchers is in line with previous research on the development of the MOFI-OTW (Multi-representation of Four-Tier Instrument on Transverse Wave) (Mufida et al., 2022).

3.2 Item Validity

Item validity test is a statistical test that is used to determine how valid a question item is to measure the variable under study. The validity of the item of a test is the accuracy of measurement

possessed by an item (which is an integral part of the test as a totality), in measuring what should be measured through the item.

In Rasch analysis, the validity value of item items (item fit) can also be determined by looking at the **Outfit MNSQ**, **Outfit ZSTD**, and **Pt-Measure CORR values**. contained in **table 10. Item: Fit Order** from the main menu Output Tables winstep program (Sumintono & Widhiarso, 2015: 71). If the item does not meet the specified criteria, it is certain that the item is not good enough so that it must be repaired or replaced (Mursidi & Soeharto, 2017). According to Boone et al, (2014) in Widhiarso and Sumintono (2015: 72), the criteria used to check the suitability of items that are not appropriate (outliers or misfits) are:

a. Received Outfit Mean Square (MNSQ) value: 0.5 < MNSQ <1.5
b. Received Outfit Z-Standard (ZSTD) value: -2.0 < ZSTD < +2.0
c. Received Point Measure Correlation (Pt Measure Corr) value: 0.4< Pt Measure Corr < 0.85

In Rasch analysis, the validity value of item items (item fit) can be determined by looking at the **Outfit MNSQ**, **Outfit ZSTD**, and **Pt-Measure CORR values**. contained in table 10. Item: Fit Order from the main menu Output Tables winstep program (Sumintono & Widhiarso, 2015: 71)

Table 5. *Item Fit Order*

Item Number	Small-scale score				Large-scale score			
	Outfit MNSQ	Outfit ZSTD	Pt-Mea Corr	Description	Outfit MNSQ	Outfit ZSTD	Pt-Mea Corr	Description
1	0.72	-1.2	0.58	Valid	0.77	-1.7	0.54	Valid
2	1.04	0.2	0.59	Valid	0.84	-1.4	0.54	Valid
3	1.05	0.3	0.43	Valid	1.15	1.3	0.43	Valid
4	0.79	-0.5	0.46	Valid	0.99	-0.1	0.52	Valid
5	1.05	0.3	0.44	Valid	0.89	-1.0	0.53	Valid
6	1.18	0.7	0.41	Valid	0.98	-0.1	0.54	Valid
7	1.04	0.2	0.51	Valid	0.86	-1.2	0.57	Valid
8	0.94	-0.1	0.52	Valid	1.07	0.7	0.43	Valid
9	1.12	0.5	0.41	Valid	1.18	1.0	0.40	Valid
10	1.24	1.1	0.41	Valid	0.96	-0.2	0.46	Valid
11	1.04	0.2	0.50	Valid	1.17	1.4	0.41	Valid
12	1.11	0.5	0.45	Valid	1.07	0.6	0.44	Valid
13	1.07	0.4	0.51	Valid	0.85	-1.1	0.54	Valid
14	0.74	-0.3	0.62	Valid	1.02	0.2	0.45	Valid
15	-0.72	-1.2	0.58	Valid	0.90	-0.7	0.51	Valid

Table 3 showed that fifteen items was developed to assess scientific literacy skills classified as Valid because they met all the criteria in the three conditions that had been set. The MNSQ outfit values obtained in the small and large scale tests were in the range of 0.67-1.18 and 0.74-1.18, respectively, meaning that the items tested were in good condition for measurement (Sumintono and Widhiarso , 2014: 128). In other words, these items can function normally in carrying out their measurements (Kurniawan & Andriyani, 2018).

The ZSTD outfit value explains the suitability of the data obtained with the model (Sumintono & Widhiarso, 2014: 128). The ZSTD outfit values obtained range from -1 to 0.7 in the small-scale test and -1.7 to 1.2 in the large-scale test. This is still included in accordance with the provisions stipulated. This range of values indicates that the data has a logical estimate (Sumintono & Widhiarso, 2014: 128). Saputri & Diana (2021) obtain a valid HOTS Problem-based instrument using the same criteria.

3.3 Reliability

Reliability is a necessary, but insufficient, criterion to assess the quality of measurement. Reliability indicates the reproducibility of the item measures if the items were administered to another sample drawn from the same population, or the reproducibility of person measures if they were tested on another occasion (Aryadoust et al., 2021).

The reliability value in the Rasch model analysis using the Winsteps program can be seen in **table 3.1** ***Summary Statistic*** on the ***Output Tables*** menu. There are three types of reliability obtained from the results of data analysis using the Rasch model, namely the reliability of the respondents (person reliability) and the reliability of the items

(item reliability) which have the criteria as contained in Table 7 as well as the Cronbach alpha value (interaction between person and item overall) with the criteria listed in Table 6 (Sumintono & Widhiarso, 2014; Fisher, 2007).

Tabel 6. *Person Reliability* and *Item Reliability* interpreting

Score	Criteria
< 0.67	Poor
0.67 – 0.80	Fair
0.81 – 0.90	Good
0.91 – 0.94	Very Good
> 0.94	Excellent

Tabel 7. *Alpha Cronbach* interpreting

Score	Criteria
< 0.5	Poor
0.51 – 0.60	Fair
0.61 – 0.70	Good
0.71 – 0.80	Very Good
> 0.80	Excellent

Tabel 8. Realiability of the test

Reliability	score	
	Small-scale	Large-scale
Person Reliability	0.74	0.73
Item Reliability	0.74	0.94
Alpha Cronbach (KR-20)	0.78	0.78

The results showed that the developed test instrument can be categorized as reliable. Table 7 shows the *results* of the reliability scores of students which tend to be stable in both small-scale and large-scale tests. In the small-scale test, a value of 0.74 was obtained, while the reliability value in the large-scale test was 0.73. Both reliability values are still included in the fair category which indicates a not very reliable person order if a similar test was *administered* to the same students (Susac et al., 2018).

Table 7 also shows the item reliability values. There is a significant difference in the value of item reliability between the results of small-scale and large-scale tests. On a small-scale test, the item reliability value is 0.74, which is categorized as fair. Meanwhile, on a large-scale test, the item reliability value increased significantly to 0.94 which is included in the very good category.

The Cronbach Alpha values shown in Table 7 are the same for small-scale and large-scale tests. Cronbach's Alpha value indicates the interaction between the person and the item as a whole. This means that the interaction between person and item in both tests is the same. This shows that the number of samples involved in the study does not have an effect on the results obtained when analyzed using the Rasch model. This has become a misunderstanding for some beginners who want to use Rasch analysis in their research (Planinic et al., 2019).

3.4 Item difficulty level

The difficulty level of the item is a measure used to indicate whether a question is included in the easy, medium or difficult category. A good question is one that is neither too easy nor too difficult. Data about the difficulty *level* of the questions can be seen in the Output Tables main menu then select Table 13. Item Measure in the winstep program (Sumintono & Widhiarso, 2015: 69). The results displayed are presented in the order of the items from the most difficult to the easiest items.

The item difficulty value is seen from the logit value of each item. At the bottom of the item measure table there is information on the standard deviation (SD) value. If the standard deviation value is combined with the average logit mean value, it can be used to categorize the level of difficulty of the items in this study. In full, can be seen in Table 9.

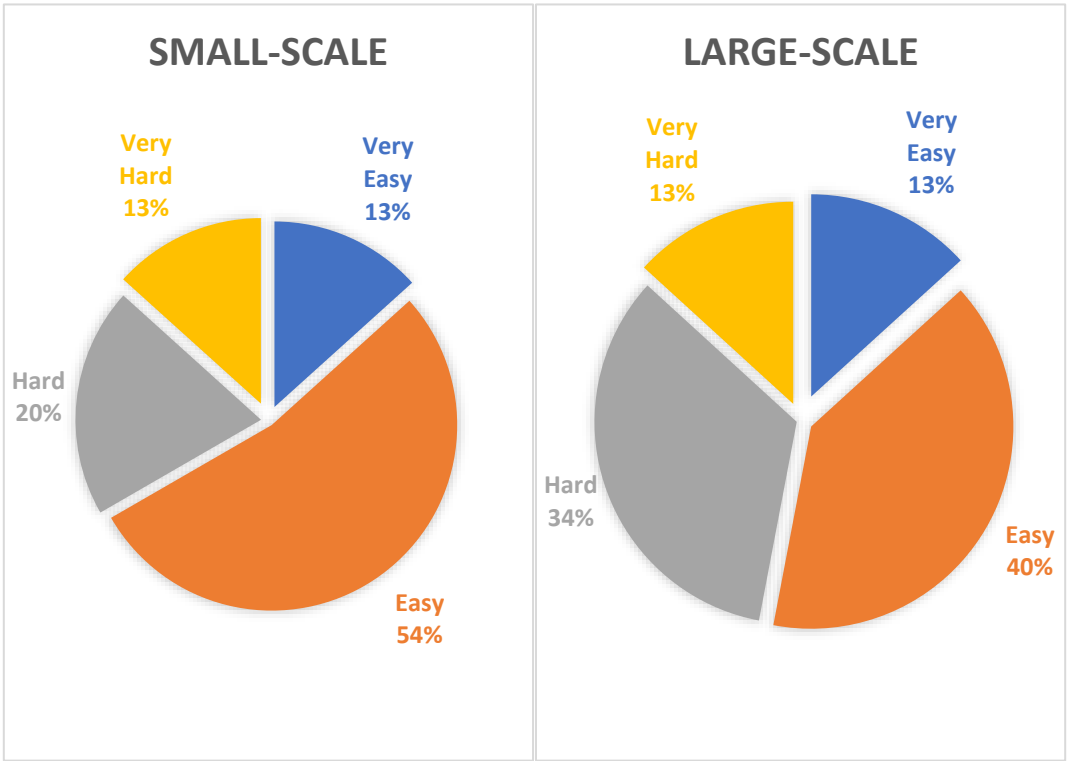
Table 9. Item difficulty level interpreting

Logit score item	Criteria
More than +1SD	Very Hard
0,0 logit + 1SD	Hard
0,0 logit – 1SD	Easy
Less than – 1SD	Very Easy

Table 10. Item Difficulty Level of test

Measure logit item in range	item difficulty level	Item code	
		Small-scale	Large-scale
Measure logit < - SD	Very Easy	N4, N6	N1, N10
-SD ≤ *measure logit* ≤ 0,00	Easy	N1, N3, N5, N7, N9, N10, N11, N13,	N2, N3, N4, N5, N6, N8
0,00 < *measure logit* ≤ SD	Hard	N2, N8, N12	N7, N11, N12, N13, N15
Measure logit > SD	Very Hard	N14, N15	N9, N14

Table 10. provides information about grouping the items based on their level of difficulty based on the results of small-scale and large-scale tests. The SD logit value becomes the measurement limit for determining each difficulty level criterion. The SD logit values obtained between small-scale and large-scale tests are not the same. In the small-scale test, the SD logit value was 0.85, while in the large-scale test, the SD logit value was 0.69. The logit SD value obtained in the small-scale test is greater than the logit SD value obtained in the large-scale test.



(a) (b)

Figure 2. Interpreting item difficulty of test (a) small-scale; (b) large-scale

Figure 2 provides an interpretation of the composition of each item difficulty level obtained from the results of small-scale and large-scale tests. There are differences in the composition of the difficulty level of items in the difficult and easy categories. Questions with an easy level of difficulty have a larger composition, namely 54% of the small-scale test results than on a large-scale test, which is only 40%. The opposite happened to questions with a difficult level of difficulty which had less composition from the results of the small-scale test than the results of the large-scale test which had a

composition of 34%. For the very easy and very difficult difficulty levels, they have the same composition, both from the small-scale and large-scale test results, which is 13%. If the questions with easy and difficult difficulty levels are included in the medium difficulty category, then the composition obtained in this research is good.

3.5 item discriminatory (ID)

Item Discriminatory (ID) is used to determine the ability of the questions to distinguish students with high abilities and low abilities (Arikunto, 2009). ID of the AKM-based numeracy assessment instrument can be determined using the Winstep software from the output table main menu, then select table 10. Item Fit Order. Point-measure correlation refers to the relationship between the level of difficulty of each individual item and the level of difficulty of the problem as a whole, where a value of one indicates a high student's ability to answer questions correctly and vice versa which shows a perfect correlation according to Rasch, a value of zero indicates no relationship between item responses and items as a whole, while a negative value indicates a problem with the item because it often gets a lower score than a high score (Smiley, 2015). The use of Point Measure Correlation (Pt. Mean Corr) can provide information about the discriminating power of a question instrument used to distinguish students' abilities.

To be able to find out the differentiating power categories of each item on the AKM-based numeracy assessment instrument in the PT-MEASURE CORR. column, Smiley (2015) provides an interpretation for each value given as shown in Table 11.

Table 11. *Point Measure Correlation* interpreting

Pt Mean Corr score	Interpreting
0.4 < *Pt. Mean Corr*	Very Good
0.30 ≤ *Pt. Mean Corr* ≤ 0.40	Good
0.20 ≤ *Pt. Mean Corr* < 0.30	Fair
Pt. Mean Corr < 0.20	Poor

Table 12. Item discriminatory result of the test

Question Number	Question Code	Small Scale		Large Scale	
		Pt Mean Corr Value	Interpretation	*Pt Mean Corr* Value	Interpretation
1	N1	0.58	Very Good	0.54	Very Good
2	N2	0.59		0.54	
3	N3	0.43		0.43	
4	N4	0.46		0.52	
5	N5	0.44		0.53	
6	N6	0.41		0.54	
7	N7	0.51		0.57	
8	N8	0.52		0.43	
9	N9	0.41		0.40	Good
10	N10	0.41		0.46	Very Good
11	N11	0.50		0.41	
12	N12	0.45		0.44	
13	N13	0.51		0.54	
14	N14	0.62		0.45	
15	N15	0.58		0.51	

Table 12 showed the results of the analysis of the differential power of the small-scale and large-scale tests of the numeracy literacy assessment instruments that have been developed. From the small-scale test, it was found that the fifteen questions tested were included in the very good category because they had a Pt value. The mean Corr is greater than 0.40 (>0.40). Different results were obtained in large-scale tests. Of the fifteen questions tested, there is one question that is included in the good category because it has a Pt value. Mean Corr 0.40. However, the developed test instrument can be categorized as feasible to use because it has a good average discriminating value so that it can distinguish which students answer correctly according to their abilities or are just lucky from guessing (Smiley, 2015).

CONCLUSION

The developed numeracy literacy assessment instrument on electromagnetic induction material AKM-based can be categorized as an appropriate instrument to use because it fulfills all the quality elements of the instrument determined from the Rasch model analysis.

ACKNOWLEDGEMENT

Thanks are addressed to the postgraduate physics education study program, Universitas Negeri Semarang, which has provided opportunities and directions during the research and to SMA Negeri 1 Pemalang who has provided the opportunity to conduct research.

REFERENCES

Aghekyan, R. (2020). Validation of the SIEVEA instrument using the Rasch analysis. *International Journal of Educational Research*, *103*(May), 101619. https://doi.org/10.1016/j.ijer.2020.101619

Arikunto, S. (2009). *Dasar-dasar Evaluasi Pendidikan (edisi revisi)*.

Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, *38*(1), 6–40. https://doi.org/10.1177/0265532220927487

Azizah, & Wahyuningsih, S. (2020). Penggunaan Model Rasch Untuk Analisis Instrumen the Use of Rasch Model for Analyzing Test. *J U P I T E K Jurnal Pendidikan Matematika*, *3*(1), 45–50.

Dewi, S. S., Hariastuti, R. M., & Utami, A. U. (2019). Analisis Tingkat Kesukaran Dan Daya Pembeda Soal Olimpiade Matematika (Omi) Tingkat Smp Tahun 2018. *Transformasi : Jurnal Pendidikan Matematika Dan Matematika*, *3*(1), 15–26. https://doi.org/10.36526/tr.v3i1.388

Himelfarb, I. (2019). A primer on standardized testing: History, measurement, classical test theory, item response theory, and equating. *Journal of Chiropractic Education*, *33*(2), 151–163. https://doi.org/10.7899/JCE-18-22

Kurniawan, U., & Andriyani, K. D. K. (2018). Analisis Soal Pilihan Ganda dengan Rasch Model. *Jurnal Statistika*, *6*(1), 34–39.

Mufida, S. N., Kaniawati, I., Samsudin, A., & Suhendi, E. (2022). Developing MOFI on Transverse Wave to Explore Students ' Misconceptions Today : Utilizing Rasch Model Analysis. *JPPIPA : Jurnal Penelitian Pendidikan IPA*, *8*(5), 2499–2507. https://doi.org/10.29303/jppipa.v8i5.2229

Mursidi, A., & Soeharto, S. (2017). an Introduction: Evaluation of Quality Assurance for Higher Educational Institutions Using Rasch Model. *JETL (Journal Of Education, Teaching and Learning)*, *1*(1), 1. https://doi.org/10.26737/jetl.v1i1.25

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 20111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Saputri, V., & Diana, H. A. (2021). Development of HOTS Problem-Based Test Instruments to Measure Level 4 Numeracy Capabilities Using Rasch Model. *Jurnal Ilmiah Mandala Education*, *7*(4), 62–66. https://doi.org/10.36312/jime.v7i4.2371

Smiley, J. (2015). Classical test theory or Rasch- A personal account from a novice user. *Shiken*, *19*(1), 16–29.

Susac, A., Planinic, M., Klemencic, D., & Milin Sipus, Z. (2018). Using the Rasch model to analyze the test of understanding of vectors. *Physical Review Physics Education Research*, *14*(2), 23101. https://doi.org/10.1103/PhysRevPhysEducRes.14.023101

Syadiah, A. N., & Hamdu, G. (2020). Analisis rasch untuk soal tes berpikir kritis pada pembelajaran STEM di sekolah dasar. *Premiere Educandum : Jurnal Pendidikan Dasar Dan Pembelajaran*, *10*(2), 138. https://doi.org/10.25273/pe.v10i2.6524

Yusuf, M., & Wulan, A. R. (2016). Penerapan Model Discovery Learning Tipe Shared Dan Webbed Untuk Meningkatkan Penguasaan Konsep Dan Kps Siswa. *Edusains*, *8*(1), 48–56. https://doi.org/10.15408/es.v8i1.1730