Available online at: http://journal.uny.ac.id/index.php/jpe

Jurnal Prima Edukasia, 11 (1), 81-95

Are the Final Semester Examination Questions for Basic Science Concepts the Quality According to the Rasch Model?

Anasufi Banawi[1] *, Muhammad Irfan Rumasoreng[2], Irawati Basta[3], Adam Latuconsina[1]
[1] Institut Agama Islam Negeri Ambon, Indonesia
[2] Yogyakarta Mercu Buana University, Indonesia
[3] Madrasah Ibtidaiyah Terpadu As-Salam Ambon, Indonesia
*Corresponding Author. E-mail: anasufibanawi@gmail.com

Received: October 3, 2022; Revised: October 27, 2022; Accepted: December 10, 2022

Abstract: Analysis of Final semester examination items in the Basic Science Concepts course using the Rasch model has not been widely carried out, even though analysing subject matter test items is a reflection of education and a noble task that educators can carry out to perfect and improve the quality of the tests they make themselves. The Rasch model is chosen so that the quality of the items does not depend on the test taker's ability to respond to answers. This study aims to describe the quality of the items from the Final Semester Examination on the Basic Science Concepts course using the Rasch model. This research used a descriptive method. Data were collected through a documented study of 10 multiple choice questions in the form of answer sheets from 42 (6 male; 36 female) students of Madrasah Ibtidaiyah Teacher Education (PGMI) at a state university in Ambon. The Rasch model, with the Winstep software program's help, is used to analyse the items. The results showed that several items, according to the Rasch model with test reliability, were in the high category. Lecturers can use existing results as diagnostic material to improve student knowledge competence and improve lectures.

Keywords: quality, items, final semester examination, basic science concepts, Rasch model

How to Cite: Banawi, A., Rumasoreng, M. I., Basta, I., & Latuconsina, A. (2023). Are the final semester examination questions for basic science concepts the quality according to the rasch model? *Jurnal Prima Edukasia, 11*(1), 81-95. doi: http://dx.doi.org/10.21831/jpe.v11i1.53633

Introduction

The lecture process has three important components: lecture objectives, learning, and evaluation of learning outcomes. All three are interrelated and influence each other. The formulation of the learning objectives strongly influences the teaching materials and methods used. Likewise, the evaluation procedure must be related to the course's material, teaching methods, and learning objectives to be achieved by students (Purwanto, 1992; Sudjana, 1991).

Students' success and achievement in certain subjects or learning can be identified through an evaluation procedure. The evaluation procedure provides an assessment of all academic aspects followed by students, including: courses, practicum, internship, comprehensive, and final examinations. The test method is a form of assessment commonly used to measure student success and achievement or certain competencies in a course (Azizah & Wahyuningsih, 2020). One form of test in universities is the final course or Final Semester Examination (*Ujian Akhir Semester*/UAS). The objectives of the examination include measuring the syllabus's effectiveness, measuring the lecture process's effectiveness, and obtaining feedback or reflection material for improving lecture activities (Jusuf et al., 2018; Kasmo, 2013; Natsir, 2010). Lecturers can obtain feedback by analysing student learning outcomes by grouping items according to the description of the concept or theme of a course (Mardapi, 2008). The final exam questions are compiled, created, used, and students' answers are corrected or checked by the lecturers (lecturers or members of the lecturer team) (IAIN Ambon, 2018).

Students of the Madrasah Ibtidaiyah Teacher Education (*Pendidikan Guru Madrasah Ibtidaiyah*/PGMI) or Elementary Teacher Education (*Pendidikan Guru Sekolah Dasar*/PGSD) study programs are prospective education graduates who have the opportunity to become teachers, including

http://dx.doi.org/10.21831/jpe.v11i1.53633

of science, at Madrasah Ibtidaiyah or Elementary Schools (MI/SD). Therefore, they must understand the concepts of science, master them, and have a correct and solid knowledge base in teaching them so that what they teach later becomes a provision of knowledge for the children to the next level of education (Sulistiawati et al., 2021). Understanding the concepts is an important aspect of achieving success in science learning. Students' cognitive understanding is indicated in their cognitive ability to understand concepts, facts, and events encountered in their own experiences through the learning they follow. A good understanding of concepts becomes a basic ability and provision to support further understanding the material. If students' understanding of concepts is low, it will affect their conceptions and provide opportunities for misconceptions to emerge (Dewi & Ibrahim, 2019).

The Basic Science Concepts course at MI/SD (GMI402) is a main or compulsory subject of the component of expertise in the field of study in the PGMI and PGSD Study Programs (Sulistiawati et al., 2021). This course is taken in the first semester (Semester 1) and contains three credits. This course discusses observation as the basis of science, quantities and units, measurements, and information related to science concepts in science teaching in MI/SD. The material discussed in the lecture meetings are Observation and Measurement, Matter and its Changes, Motion and Energy, Temperature and Heat, Living Things and Their Environment, Solar System, Mid-Semester Examination (*Ujian Tengah Semester*/UTS), Water, Air, Sound, Rock and Soil, Electricity, and Magnetism, Light, and Simple Chemistry, and after that the Final Semester Examination (*Ujian Akhir Semester*/UAS). The mastery of the materials and correct understanding of concepts are important aspects of knowledge, in addition to attitudes and skills, which are the course's objectives. Students of PGMI/PGSD must possess the aspects when taking the Basic Science Concepts course (Banawi, 2021). One of how students' understanding of concepts can be known is through item analysis of the examination material.

Assessing tests made by educators (lecturers or teachers) through item analysis, honestly reviewing the prepared questions, and checking the validity of items and test reliability is a noble and commendable effort that educators can do because not all educators do these tasks. Moreover, some educators find it difficult to realise that the test questions they have prepared are still invalid or imperfect according to certain rules (Al Fajar et al., 2022; Arikunto, 2006). As mentioned earlier, the lecturers (or lecturer team) make the final examination questions, and one of the examination objectives is to obtain feedback or reflection material for lecture improvements. To realise this in lectures, the lecturers (or teams of lecturers) need to assess the course examination questions through item analysis activities, including the Basic Science Concepts course in the PGMI and PGSD Study Programs. Items can be improved in quality through item analysis activities; therefore, these activities are important (Safari, 2003, p. 64). The characteristics of quality test items and test kits contribute to the quality of the test itself and the quality of the assessment.

There are ways to analyse items, both qualitatively (material, form/construction, language), empirically, and quantitatively. Quantitatively, it can be done through traditional analysis or Classical Test Theory (CTT) and Item Response Theory (IRT) (Jackson et al., 2002; MacCann & Stanley, 2006; Anggraini & Suyata, 2014; Tabatabaee-Yazdi et al., 2018; Erfan et al., 2020). CTT has several limitations, including: (1) the statistics of test items (difficulty level, differentiating power of questions) are always influenced by the characteristics of the participants (test takers). The difficulty level of the questions will be high if the participants' abilities are low. The homogeneity of the participants largely determines the differentiating power of the questions (biserial point correlation coefficient); (2) the estimation of the participants' ability is influenced by the test items; (3) the estimated error score covers all participants; (4) there is no information on the response of each participant to the item; and (5) the reliability estimation uses elusive parallel assumptions. The use of IRT can cover these weaknesses. One form of IRT, item analysis, uses the Rasch Model or IRT with one logistic parameter model (1PL) (Alfarisa & Purnama, 2019; Mardapi, 2008). Winstep Software, a computational tool, is suitable for using the Rasch model to calculate item parameters (Azizah & Wahyuningsih, 2020; Muslihin et al., 2022). In this study, some parameters of the Basic Science Concepts Course Final Examination were analysed with the Rasch model using Winstep Software.

Research on the Basic Science Concepts course is not new. There have been similar studies such on understanding concepts (Dewi & Ibrahim, 2019), increasing understanding of concepts (Sulistiawati et al., 2021), analysis of learning difficulties (Surya & Marta, 2017; Winarti, 2021), analysis of student abilities in solving problems (Jusuf, 2015), and analysis of items with the Anates Program (Alpusari, 2015). There are also several studies on the Rasch model, including on the validity of the e-module with the Rasch model (Ramadhani & Fitri, 2020), analysis of problem-solving abilities with the Rasch model (Dwinata, 2019), the Rasch model to analyse misconceptions (Khong & Lim, 2019), the use of the Rasch model for analysis of test instruments (Alfarisa & Purnama, 2019; Azizah & Wahyuningsih, 2020; Purba, 2018; Herwin, Tenriawaru, & Fane, 2019), quality analysis of series and parallel questions (Erfan et al., 2020), and an analysis of Higher Order Thinking Skill (HOTS) test questions (Al Fajar et al., 2022; Rochman & Hartoyo, 2018). However, the novelty of this current study is the analysis of the quality of the final exam questions for the Basic Science Concepts course (GMI402) using the Rasch model assisted by Winstep. Studies from previous research are used as initial and complementary instructions in conducting this study and delivering the discussions.

However, it is ironic that examples of analysis results of final examination questions in universities are still minimal. In addition, the analysis of the final exam questions for the Basic Science Concepts course using the Rasch model has not been carried out. Therefore, this study is significant. Referring to the description above, the main problems discussed in this study are: "How is the quality of the Final Semester Examination questions in the Basic Science Concepts course made by the lecturer based on the Rasch model?" In line with the existing problems, this study aims to describe the quality of the items from the Final Semester Examination for the Basic Science Concepts course using the Rasch model. The Rasch model is chosen so that the quality of the items does not depend on the test taker's ability to respond to answers. The good quality of the exam questions can improve the measuring function and accurate test decisions (Herwin, Tenriawaru & Fane, 2019). It is hoped that the results of this study can provide information on the characteristics of the questions and can be used as material to maintain and improve the quality of final exam questions for the Basic Science Concepts course, as a diagnostic material to achieve student knowledge competencies to improve the course in the PGMI Study Program in the future, and as a reference for future researchers who will carry out similar activities in certain courses and or other related research.

Methods

The descriptive method was used in this study. The research subjects were students enrolled in the Basic Science Concepts course in the PGMI Study Program at a state university in Ambon City in the Odd Semester of the 2021/2022 Academic Year. A total of 42 students (6 males; 36 females) were selected as subjects by purposive sampling or with certain considerations (Sugiyono, 2013, p. 124). The object of the study was ten multiple-choice questions for the Semester's Final Examination of the Basic Science Concepts, which were self-made or prepared by the lecturer in charge of the course (Arikunto, 2006, p. 204). The focus of this study was analysing the final examination questions. Table 1 summarises the final exam questions for the Basic Science Concepts Course.

Table 1. Summary of multiple choices instruments of basic science concepts course final examination

No.	Question Indicator	Cognitive Level	Item Number	Quantity of Items	Answer Key
1	Students understand the concept of water.	C1	1	1	B
2	Students understand the concept of air.	C2	2	1	E
3	Students understand the concept of sound.	C3	3	1	D
4	Students understand the concept of rocks and soil.	C2, C4	4, 5	2	A, A
5	Students understand the concept of electricity and magnets.	C2, C1	6, 7	2	C, B
6	Students understand the concept of light.	C5	8	1	C
7	Students understand the concept of simple chemistry.	C1, C4	9, 10	2	A, C
	Total			10	

The data were collected through document studies on student answer sheets (Abdillah, 2013). The data were analysed from the multiple choice questions using the Rasch IRT model assisted by the Winsteps and SPSS software (Erfan et al., 2020; Ghozali, 2006). The analysis included testing the assumptions of item response theory, namely: unidimensionality, parameter invariation, and local independence (Alfarisa & Purnama, 2019; Hambleton et al., 1991, p. 9; Retnawati, 2014, p. 1). The unidimensional assumption was met through factor analysis using the K-M-O (Kaiser-Meyer-Olkin) Test and Bartlett's Test. The assumption of local independence was fulfilled as the unidimensional assumption was fulfilled. The parameter invariance assumption test was done by looking at the correlation between the parameter invariation of the participant's ability and the parameter in a variation of item difficulty level. If the correlation coefficient was positive and high, then the parameter invariance assumption was fulfilled (Alfarisa & Purnama, 2019; Retnawati, 2014). The quality of the items includes (1) the validity of the items with (Column) Item with the acceptance criteria of Outfit Mean Square (MNSQ) ranging from 0.5<MNSQ<1.5; Outfit Z-Standard (ZSTD) value between -2<ZSTD<+2; and the Point Measure Correlation (Pt. Mean Corr) value at 0.4<Pt. Mean Corr<.85 (Azizah & Wahyuningsih, 2020; Dwinata, 2019; Khong & Lim, 2019; Purba, 2018; Rochman & Hartoyo, 2018; Ramdani et al., 2020). Construct validity related to the ability to measure all subjects (range variables) referred to item dimensionality with the Raw variance explained by measure criteria being more than 20% (Lestari et al., 2020); (2) the instrument reliability by looking at summary statistics; (3) the item difficulty was known from item dimension and item measure; and (4) the perception or bias by taking into account the Differential Item Functional (DIF) plot and based on the Winstep output. If the p-value<0.05, the bias was significant (Ramadhani & Fitri, 2020; Sumintono & Widhiarso, 2015). Some of the characteristics of the items were adjusted to the Rasch Model on the instrument quality criteria rating scale (Fisher Jr., 2007; Meyer & Hailey, 2012). The quality of the items from the explanation above was summarised in Table 2.

Table 2. Summary of question quality using the rasch model

No.	Aspect	Indicator	Indicator Element	Criteria
1	Item validity	Item (column): fit order item fits the model if it fulfils one or both requirements	Outfit MNSQ Outfit ZSTD and Pt. Meas. Corr.	0.5 < MNSQ < 1.5 −2 < ZSTD < +2 0.4<Pt. Mean Corr<0.85
		dimensionality item is related to construct validity, can find out *range variable*	Raw variance explained by measure	Raw variance explained by measure > 20%
		The existence of contaminant factor	Raw variance explained by measure	< 50% = Poor 50 - 60% = Fair 60 - 70% = Good 70 - 80% = Very Good > 80% = Excellent
			Unexplained variance in contrast 1-5 PCA of residual	< 3% = Excellent 3 - 5% = Very Good 5 - 10% = Good 10 - 15% = Fair > 15% = Poor
2	Instrument reliability	Summary statistics	Item reliability Person reliability	Reliability value (person and item); > 0.94 = Excellent 0.91 – 0.94 = Very good 0.81 – 0.90 = Good 0.67 – 0.80 = Fair < 0.67 = Poor
3	Question difficulty	Wright Map or item measure, item fit (person-item map)	Measure value (logit)	Measure value (logit): Measure logit<-SD logit = item very easy -SD logit≤measure logit ≤ 0.00 = item easy 0≤ measure logit ≤ SD logit = item difficult Measure logit>SD logit = item very difficult

		Differentiating power from individual ability	Pearson measures person fit. Differentiating value (H).	Grouping according to the separation index, $H = \frac{[(4x\text{separation}) + 1]}{3}$
4	Bias or perception	Differential Item Functioning (DIF) plot	Winstep DIF output table	Difference of perception significant at p<0.05

Source: Processed Data

Results and Discussion

Using the Rasch model, the characteristics of the items are directed at the difficulty level parameter. The difficulty level is part of the quality of the items in the semester final examination of the Science basic concepts course (UAS MK-KDIPA) in addition to validity, reliability and bias. Before the item analysis of the Basic Science Concepts Final Examination, the item response theory assumption was conducted. Assumption test includes unidimensional parameters invariation and local independence. The output of the SPSS software is presented in Table 3.

Table 3. KMO and Bartlett's test[a]

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.513
Bartlett's Test of Sphericity	Approx. Chi-Square	55.887
	Df	45
	Sig.	.128

a. Based on correlations

The unidimensional assumption was tested by factor analysis using K-M-O and Bartlett's Test. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO-MSA) value was 0 to 1. To conduct the factor analysis, the K-M-O value had to be greater than 0.5 (Ghozali, 2006, p. 49). The KMO test was related to the adequacy of the sample, while Bartlett's Test of Sphericity was related to the fulfilment of homogeneity (Setiawan, Fajaruddin & Andini, 2019; Hartono et al., 2022). Based on the output obtained, the KMO value was 0.51> 0.50, and then the analysis was continued because it met the adequacy of the sample.

Table 4. Total variance explained

Component	**Initial Eigenvalues[a]**		
	Total	**% of Variance**	**Cumulative %**
1	.484	22.670	22.670
2	.332	15.541	38.211
3	.314	14.697	52.908
4	.236	11.060	63.968
5	.208	9.722	73.690
6	.196	9.189	82.880
7	.156	7.326	90.206
8	.101	4.744	94.950
9	.061	2.854	97.805
10	.047	2.195	100.000

Table 4 shows no Eigen value is greater than one. The largest eigenvalue is in Factor-1, which is 0.484, so this factor is more dominant than other factors. Factor 1 can explain 22.67% of the variation. It can be stated that the final test questions were unidimensional. The fulfilment of the local independence assumption is in line with the unidimensional assumption. Local independence is detected by fulfilling the unidimensional assumption (Retnawati, 2014, p. 3).

The next assumption test was the parameter invariation. This assumption could be proven by a large or high correlation between the sample groups and the calibration results of the item parameters (Widhiarso, 2011). The correlation between variables could also be known from the partial correlation between variables, assuming a constant or fixed value in other variables. In SPSS, the partial correlation was shown by the anti-image correlation matrix as a negative partial correlation value. The value of the correlation coefficient (*r*) category was sufficient in the range of 0.4 to 0.6. Values of *r* above that interval or close to 1 are categorised as high and very high (Arikunto, 2006, p. 75). The results of the anti-image matrix are shown in Table 5. Based on the table, the correlation value was indicated in the high and medium categories.

Table 5. Results of anti-image matrices

Anti-image Correlation	S1	.337[a]	.067	-.072	.030	.135	-.271	.329	-.280	.040	-.119
	S2	.067	.759[a]	-.047	-.116	.056	-.090	-.087	.006	.028	-.015
	S3	-.072	-.047	.562[a]	.108	.063	-.080	.010	-.023	.164	-.375
	S4	.030	-.116	.108	.650[a]	.104	-.074	-.180	.034	.171	-.110
	S5	.135	.056	.063	.104	.467[a]	-.118	.048	-.232	-.243	.071
	S6	-.271	-.090	-.080	-.074	-.118	.660[a]	-.211	-.072	-.114	-.097
	S7	.329	-.087	.010	-.180	.048	-.211	.502[a]	-.515	-.014	-.284
	S8	-.280	.006	-.023	.034	-.232	-.072	-.515	.463[a]	.269	.253
	S9	.040	.028	.164	.171	-.243	-.114	-.014	.269	.481[a]	-.205
	S10	-.119	-.015	-.375	-.110	.071	-.097	-.284	.253	-.205	.498[a]

a. Measures of Sampling Adequacy (MSA)

The results above show that unidimensionality, parameter invariation, and local independence have been met. Furthermore, the results of the item analysis (output) using Winstep were directed to answer research problems related to the quality of the items, including the difficulty level, item validity, test reliability, and bias.

Table 6. Results of item analysis of basic science concepts course final examination

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODELS. E.	INFIT		OUTFIT		PTMEASURE-AL			EXACT MATCH		Item
					MNSQ	ZSTD	MNSQ	ZSTD		CORR.	EXP.	OBS%	EXP%	
9	4	42	1.64	0.54	1.03	0.21	3.13	2.39	A	0.06	0.21	89.7	89.8	S9
2	1	42	3.15	1.02	1.04	0.36	1.26	0.62	B	0.04	0.10	97.4	97.5	S2
4	22	42	-1.08	0.35	1.10	0.80	1.15	0.96	C	0.39	0.47	64.1	68.0	S4
1	16	42	-0.35	0.35	1.12	0.98	1.13	0.73	D	0.32	0.40	61.5	66.5	S1
5	18	42	-0.59	0.35	1.06	0.58	1.03	0.26	E	0.39	0.42	61.5	65.6	S5
3	14	42	-0.10	0.36	1.04	0.36	1.01	0.13	e	0.35	0.38	66.7	67.6	S3
10	4	42	1.64	0.54	0.97	0.05	0.81	-0.11	d	0.24	0.21	89.7	89.8	S10
6	28	42	-1,87	0.39	0.89	-0.52	0.81	-0.69	c	0.60	0.52	76.9	74.9	S6
7	27	42	-1.73	0.38	0.88	-0.64	0.80	-0.86	b	0.60	0.52	74.4	73.7	S7
8	19	42	-0.71	0.35	0.80	-1.83	0.77	-1.54	a	0.58	0.44	79.5	65.1	S8
MEAN	15.3	42.0	0.00	0.46	0.99	0.04	1.19	0.19				76.2	75.8	
P.SD	9.1	0.0	1.55	0.20	0.10	0.79	0.67	1.04				12.2	11.4	

From the output of Table 6 and confirmed by Table 2, it was found that the questions (items) were in the Outfit Mean Square interval; Outfit 0.5<MNSQ<1.5 (9 items: S1, S2, S3, S4, S5, S6, S7, S8, and S10) and one item did not fit (valid) namely item Number-9, S9 (3.13>1.5); Items in the Outfit Z-Standard interval; Outfit -2<ZSTD<+2 (9 items: S1, S2, S3, S4, S5, S6, S7, S8, and S10) and one item did not fit (invalid), namely Number-9, S9 (+2.39>+2). The largest ZSTD value was +2.39. The sample size can affect the ZSTD value. Generally, large samples have a ZSTD value above 3 (Ramadhani & Fitri, 2020). Items in the Point Measure Correlation interval; 0.4<Pt.Meas.Corr.<0.85 (3 items: S6, S7, and S8). Meanwhile, seven items were not appropriate, namely: S9 (0.06<0.4), S2 (0.04<0.4), S4 (0.39<0.4), S1 (0.32<0.4), S5 (0.39<0.4), S3 (0.35<0.4), and S10 (0.24<0.4). The item limits are said to be fit (valid) if they meet one or both of the requirements for Outfit Mean Square, Outfit Z-Standard, and Point Measure Correlation (Azizah & Wahyuningsih, 2020). Therefore, it was concluded that from the ten questions (items) of the Basic Science Concepts Course Final Examination, there was one invalid item (S9) and nine useful items (S1, S2, S3, S4, S5, S6, S7, S8, and S10). The items that met the valid criteria could be used as part of the test instrument. The next step was the item dimensionality analysis related to the construct validity of the Final Examination questions by looking at the value of natural variance explained by measures and matching it with the criteria (Table 2).

Table 7. Results of the item dimensionality test

Table of STANDARDISED RESIDUAL variance in Eigenvalue units = Item Information units					
		Eigenvalue	**Observed**		**Expected**
Total raw variance in observations	=	15.0423	100.0%		100.0%
Raw variance explained by measures	=	5.0423	33.5%		32.9%
Raw variance explained by persons	=	1.0656	7.1%		7.0%
Raw variance explained by items	=	3.9767	26.4%		26.0%
Raw unexplained variance (total)	=	10.0000	66.5%	100.0%	67.1%
Unexplained variance in 1st contrast	=	1.7234	11.5%	17.2%	

The values in Table 7 were confirmed by Table 2 and the Raw variance explained by measures was 32.9%>20%. Thus, the Basic Science Concepts Final Examination instrument had construct validity and could measure all students' abilities. The value of the Row variance explained was 32.9%<50% at the Poor level, and the Unexplained variance in the first contrast was 17.2%>15% at the Poor level. It was concluded that in the scale, the items were influenced by other factors that were not measured in the question or that there were still contaminant factors from the questions given. After knowing the item's validity, the next analysis was the instrument's reliability.

Table 8. Results of summary statistics

Person	**42 INPUT**		**42 MEASURED**		**INFIT**		**OUTFIT**	
	TOTAL	**COUNT**	**MEASURE**	**REAL SE**	**IMNSQ**	**ZSTD**	**OMNSQ**	**ZSTD**
MEAN	3.6	10.0	-1.04	0.92	1.00	0.0	1.19	0.2
P. SD	1.7	0.0	1.23	0.29	0.33	1.0	1.59	0.9
REAL RMSE 0.97		TRUE SD 0.75		SEPARATION 0.78		Person RELIABILITY 0.38		
Item	10 INPUT		10 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REAL SE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	15.3	42.0	0.00	0.47	0.99	0.0	1.19	0.2
P. SD	9.1	0.0	1.55	0.20	0.10	0.8	0.67	1.0
REAL RMSE 0.51		TRUE SD 1.46		SEPARATION 2.85		Item RELIABILITY 0.89		

From the Winstep output (Table 8), there were 42 people and ten questions. The Person Measure value of -1.04 logit<0.0 logit showed the tendency that the student's abilities were smaller than the students' difficulty level or the average student's ability to answer questions was smaller than the level of difficulty of the questions. The item reliability value showed 0.89, which included the Good Category, while the Person reliability was at 0.38 (Weak category). It shows that students' consistency in answering was still weak, while the quality of the questions in the instrument was good. Thus, lecturers or other researchers could use the Basic Science Concepts Final Examination questions based on the reliability aspect. After analysing the reliability of the test and the validity of the items, the next step was to analyse the difficulty of the items.

The difficulty level of the Basic Science Concepts Final Examination questions can be seen by looking at the Wright Map, as shown in Figure 1 below.

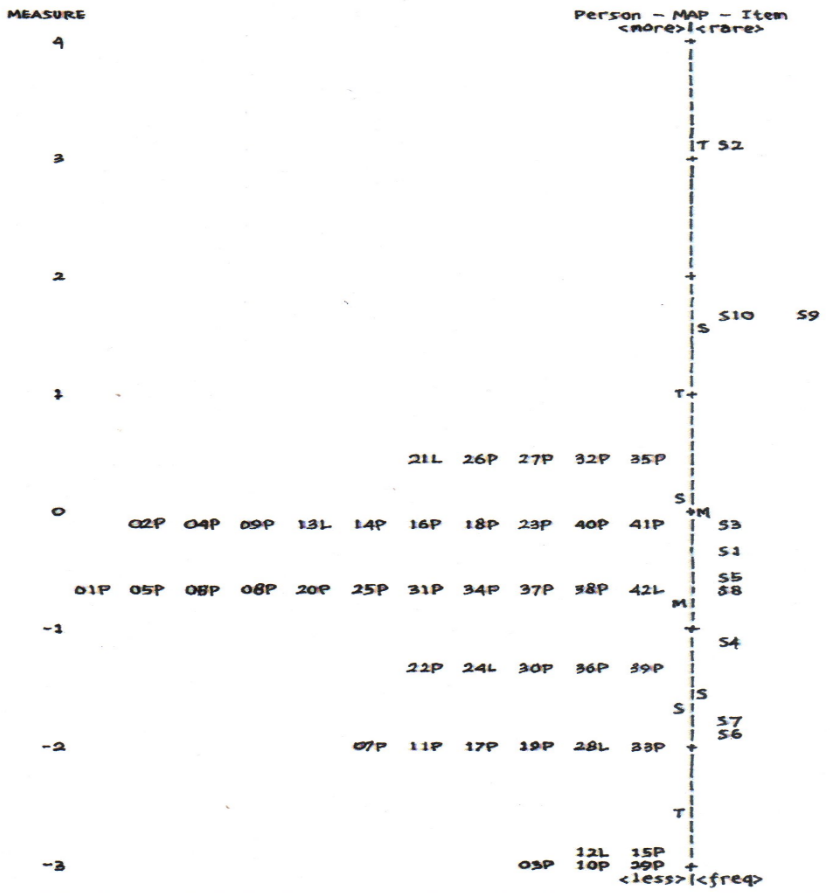

Figure 1. Wright map of basic science concepts final examination

From the Wright Map analysis (in Figure 1), it was found that three questions had a level of difficulty above the student's ability to answer them, namely questions with codes S2, S10 and S9, and students having the lowest ability to answer these questions were students with codes 12L, 15P, 03P, 10P, and 29P (4 Female and 1 Male).

Based on Table 6 and Table 8, it is known that the P.SD or SD logit value was 1.55. When this value was associated with Table 2, a measure (logit) value was obtained, showing the items' difficulty levels and students' abilities. The results were as follows: 2 items were in the Very Easy category, namely questions with codes S6 (-1.87) and S7 (-1.73); 5 items were in the Easy category, namely questions with codes S4 (-1.08), S1 (-0.35), S5 (-0.59), S3 (-0.10), and S8 (-0.71); and three items were in the Very Difficult category, namely questions with codes S9 (1.64), S2 (3.15), and S10 (1.64). The average measure (logit) was 0.00, and the difficulty level of the questions was Medium (Boone, Staver & Yale, 2014). Logit or log odds unit is the participants' ability to solve questions, which is influenced by the participant's level of ability and the items' difficulty (Englehard, 2013). The measured value (logit) is an indicator of the level of difficulty of the items (Erfan et al., 2020).

From Table 8, it can be seen that the Person Separation score was 0.78. It means that the items were rather not sensitive in accommodating student abilities. However, an Item Separation of 2.85 (close to 3) indicated that the participants were varied enough to detect the items given. From the Item Separation value, it was obtained that H = 4.13 and rounded up to 4, so there were four groups of items. The Basic Science Concepts Final Examination questions instrument was good because the H value was 4. The number of item groups according to the H value, namely: difficult, medium, easy, and very easy. The greater the strata (H) value, the better the instrument because it can identify groups (respondents and items) in more detail (Purba, 2018).

The information function can express the strength of an item in the test to reveal the latent trait (Alfarisa & Purnama, 2019). The next step was detecting conformity with the Rasch Model. The corresponding Winstep output is shown in Table 9 below.

Table 9. Output from Guttman scalogram of basic science concepts final examination

GUTTMAN SCALOGRAM OF RESPONSES:											
Person	Item										Person
	6	7	4	8	5	1	3	9	10	2	
21	1	1	1	1	0	1	0	0	1	0	21L
26	1	1	1	1	0	1	1	0	0	0	26P
27	1	1	0	0	1	1	0	1	1	0	27P
32	1	1	1	1	1	0	1	0	0	0	32P
35	1	1	1	0	1	0	1	1	0	0	35P
2	1	1	1	1	1	0	0	0	0	0	02P
4	1	1	1	1	0	1	0	0	0	0	04P
9	1	1	0	1	1	0	1	0	0	0	09P
13	1	1	1	1	0	1	0	0	0	0	13L
14	1	1	0	1	1	1	0	0	0	0	14P
16	1	1	0	1	1	0	1	0	0	0	16P
18	1	1	1	1	1	0	0	0	0	0	18P
23	0	1	1	1	1	1	0	0	0	0	23P
40	1	0	1	1	0	1	1	0	0	0	40P
41	1	1	1	1	0	0	1	0	0	0	41P
1	0	1	1	0	1	0	1	0	0	0	01P
5	1	0	1	0	0	1	1	0	0	0	05P
6	1	1	1	0	0	0	0	0	1	0	06P
8	1	1	1	0	0	0	0	0	1	0	08P
20	1	1	1	0	0	1	0	0	0	0	20P
25	0	1	0	1	0	0	1	0	0	1	25P
31	1	1	1	1	0	0	0	0	0	0	31P
34	1	0	0	0	1	1	1	0	0	0	34P
37	1	1	0	1	1	0	0	0	0	0	37P
38	0	0	0	1	1	1	1	0	0	0	38P
42	1	1	0	1	1	0	0	0	0	0	42L
22	1	1	1	0	0	0	0	0	0	0	22P
24	0	1	0	1	0	1	0	0	0	0	24L
30	1	0	0	0	1	1	0	0	0	0	30P
36	1	0	0	0	1	0	0	1	0	0	36P
39	0	0	1	0	1	0	1	0	0	0	39P
7	1	0	0	0	0	1	0	0	0	0	07P
11	0	0	1	0	1	0	0	0	0	0	11P
17	1	0	0	0	0	1	0	0	0	0	17P
19	0	1	1	0	0	0	0	0	0	0	19P
28	1	1	0	0	0	0	0	0	0	0	28L
33	0	1	0	0	0	0	1	0	0	0	33P
12	0	0	0	0	0	0	0	1	0	0	12L
15	0	0	1	0	0	0	0	0	0	0	15P
3	0	0	0	0	0	0	0	0	0	0	03P
10	0	0	0	0	0	0	0	0	0	0	10P
29	0	0	0	0	0	0	0	0	0	0	29P

Table 9 shows that carelessness occurred in some students who were able to answer difficult questions but could not answer the easy questions correctly. Students who experienced this were students with codes 21L, 26P, 27P, 32P, 35P, 04P, 9P, 13L, 14P, 16P, 23P, 40P, 41P, 01P, 05P, 06P, 08P, 20P, 25P, 34P, 37P, 38P, 42L, 24L, 30P, 36P, 39P, 07P, 11P, 17P, 19P, 33P, 15P, and 12L. Students who were indicated as making a lucky guess in answering the Basic Science Concepts Final Examination questions were students with codes 01P, 25P, and 38P. There were two indications of students cheating, namely with codes 06P and 08P.

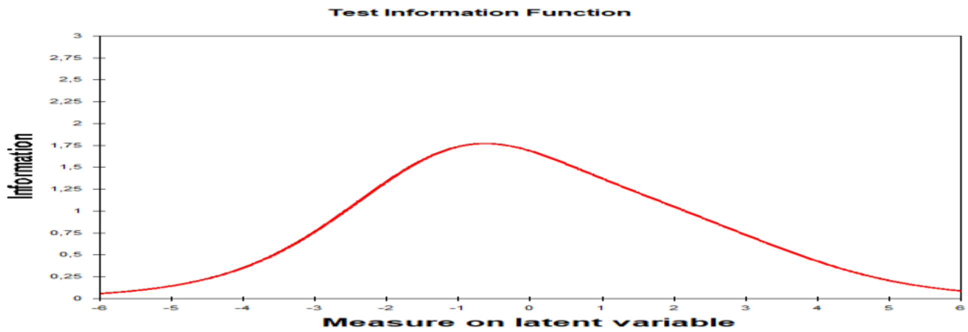

Figure 2. The function of information in the basic science concepts final examination

Figure 2 presents the information function of the Basic Science Concepts Final Examination, where the Y-axis is the value of the function, and the X-axis is the student's ability level. The information function shows that the questions suit students with moderate-level abilities.

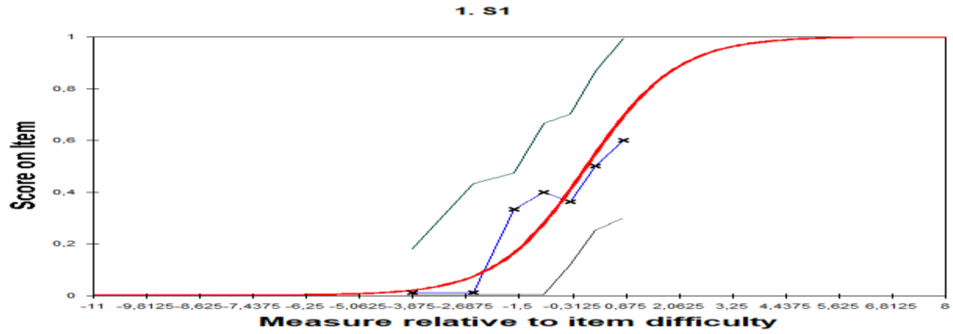

Figure 3. Relative measure of item difficulty of basic science concepts final examination

Figure 3 shows that the higher a person's ability, the higher the ability to answer questions correctly. The graph of the questions follows the Rasch Model. Ability relates to the ability possessed by the test taker. The test takers with better (higher) abilities are very likely to answer questions correctly, and conversely, low abilities cause individuals to have a small chance of answering questions correctly (Erfan et al., 2020; Purba, 2018; Sumintono & Widhiarso, 2015, p. 3). The analysis related to the quality of the next item is related to bias or perception.

Testing for bias in this study was based on aspects of gender (female and male) or demographics (Ramadhani & Fitri, 2020). The Differential Item Functional (DIF) test results are shown in Table 10 below.

Table 10. The output of DIF of Basic science concepts final examination

DIF class/group specification is: DIF=$S3W1

Person	SUMMARY DIF			BETWEEN-CLASS/GROUP		Item	
CLASSES	CHI-SQUARED	D.F.	PROB.	UNWTD MNSQ	ZSTD	Number	Name
2	0.6153	1	0.4328	0.7189	0.25	1	S1
2	0.1848	1	0.6673	0.1708	-0.47	2	S2
2	1.7807	1	0.1821	3.9409	1.70	3	S3
2	0.8394	1	0.3596	1.0085	0.48	4	S4
2	1.5563	1	0,2122	2.1092	1.07	5	S5
2	0.0078	1	0.9296	0.0105	-1.18	6	S6
2	1.3226	1	0.2501	1.6995	0.88	7	S7
2	1.5533	1	0.2126	2.0226	1.03	8	S8
2	0.4135	1	0,5202	0.4957	0.03	9	S9
2	0.4135	1	0.5202	0.4957	0.03	10	S10

From Table 10, it appears that the probability of all items was above 0.05 or Prob>0.05, which meant that there was no significant difference in perception between the two genders (male and female). All students had the same (significant) perception of the 10 Basic Science Concepts Final Examination questions.

Using the Rasch model in relation to item analysis is not new. There have been several previous studies (Purba, 2018; Alfarisa & Purnama, 2019; Herwin, Tenriawaru, & Fane, 2019; Azizah & Wahyuningsih, 2020; Nur et al., 2020; Ilfiandra et al., 2022; Astuti et al., 2022). However, this study is different from the existing ones because this research relates the results of the analysis of exam items in the Basic Science Concepts course (GMI402) through the Winstep-assisted Rasch model with the quality of these items.

This study found that of the ten items of the Basic Science Concepts Final Examination, nine valid items and one invalid item (S9). Many items were in line with the Rasch model's test reliability (0.89) in the good or high category. The average difficulty level of the items was in the medium category, so it was suitable to be tested on male and female students at that level of ability. The items that were not valid would be improved. The improvements were according to the validity criteria of the Rasch model (Rasch Model) and the stages of analysis (Ramadhani & Fitri, 2020).

A lecturer needs to analyse test items. It is important that items that match the criteria can be maintained and those that are not suitable can be improved so that the Basic Science Concepts Final Examination questions contain quality that lecturers or other researchers can use. This study's results align with previous research (Alfarisa & Purnama, 2019; Azizah & Wahyuningsih, 2020; Purba, 2018) that the moderate level of difficulty was very suitable for use in groups of students with moderate ability levels. Therefore, the optimal Basic Science Concepts Final Examination questions were tested on groups of students with moderate abilities.

The item analysis results can describe students' abilities (Banawi et al., 2022; Kumaidi, 1999). Students' achievement on the ten Basic Science Concepts Final Examination questions tested provided an overview of students' understanding of concepts in the Basic Science Concepts course in the PGMI Study Program. According to the Wright Map analysis, Question Number-2 (S2), which was related to the concept of air, and Number-9 (S9), as well as Number-10 (S10), which were related to simple chemistry concepts, were the questions that students had difficulty answering. Lecturers can use these results to improve the test questions used in the Basic Science Concepts course final examination and improve lectures. Question Number-2 (S2) and Question Number-9, according to Table 1, were questions with low cognitive levels; C2 (understanding) and C1 (memory). According to cognitive theory, questions with the C1 qualification measure the lowest cognitive level (Rochman & Hartoyo, 2018). While question Number 10 was at a moderate cognitive level, C4 (analysis). However, many students faced difficulties in answering these questions. In theory, all examinees should have answered easy questions correctly, but this was not the case. Likewise, in the lectures in class, the students felt they did not experience difficulties in learning. However, when the test was carried out, their scores were unsatisfactory. These results indicated that the students had learning difficulties (Ma'rifah, 2017). It was strongly suspected that one of the factors causing students' learning difficulties in the Basic Science Concepts course was related to the characteristics of the science lecture material which required students to be skilled in applying certain concepts in the form of practicum (Winarti, 2021). Lecturers can only analyse the suitability between the questions used and the lesson plans that have been made previously (Semester Learning Plan or Lecture Program Unit) (Alpusari, 2015; Nurharyanto & Retnawati, 2020). It was a self-reflection so that the lecturers of Basic Science Concepts could improve their lectures (Sulistiawati et al., 2021; Yusron & Sudiyatno, 2021; Black & Wiliam, 2018).

The lecturer's efforts to analyse the compiled test items are a reflection of education and a noble task and need to be replicated by other educators (lecturers or teachers) to perfect and improve the quality of the tests (Arikunto, 2006; Prasetyo, 2017; Syadiah & Hamdu, 2020). The item analysis results are a source of information to other parties about the achievement of student knowledge competencies and about the materials that the students have studied and certain concepts, they have not mastered. Understanding concepts (Basic Concepts of Science) is very important for PGMI and PGSD students because understanding these concepts provides knowledge to support subsequent understanding courses such as Science Learning in MI/SD. A good understanding of concepts becomes the basis for supporting further understanding of the material (Fatmawati, 2016; Dewi & Ibrahim, 2019). In addition, the Basic Science Concepts course supports PGMI and PGSD students as prospective classroom teachers (Surya & Marta, 2017). The results of this study are expected to help educators to conduct an objective self-assessment of the tests they have prepared and diagnose students' learning difficulties. Objective educational reflection needs to be done and is also an important part or contribution of this research for readers in general and educators and students in particular.

Conclusion

The quality of the ten items of the test can be seen in the difficulty level, validity, test reliability, and bias. A number of these items are in line with the Rasch model with good or high category test reliability. There are two questions (S6, S7) in Very Easy category, five questions (S1, S3, S4, S5, S8) in Easy category, and three questions (S2, S9, S10) in Very Difficult category. The average difficulty level of the items is Medium. The questions will function optimally or are suitable for testing on male and female students with medium-level abilities: nine (90%) valid items and one (S9) invalid item. Item reliability value indicates a good category, while person reliability value indicates a weak category. Thus, the questions can be used by lecturers or other researchers based on the reliability aspect. Regarding bias, there was no significant difference in perception between men and women.

This study implies that lecturers need to make improvements to test questions and improve lectures. In addition to having advantages, this study has some limitations. They are related to the research method used, including the absence of treatment on the subject because this study only focused on analysing test questions and students' answers, and the number of questions and test takers was relatively small. It is recommended that further research develop multiple choice questions in the Basic Science Concepts course final examination by adding more items. In addition, the examination questions for this course or other courses can add items in the form of short answers accompanied by a scoring rubric on each item.

References

Abdillah. (2013). Profile of errors in solving one and two variable linear equations and inequalities of second semester students in Class of 2012/2013 majoring in Mathematics Education IAIN Ambon. *Horizon Pendidikan*, *8*(1), 1–13. http://repository.iainambon.ac.id/269/1/HORIZON%20PENDIDIKAN%20%3B%282013%29%20Mata%20Pelajaran%20IPA-FISIKA%20dan%20Pembentukan%20Karakter%20Peserta%20Didik..PDF

Al Fajar, S. N., Sitompul, S. S., & Arsyid, S. B. (2022). Analysis of students' ability to complete the higher order thinking skill (hots) test with the help of the winstep application on simple machine material. *Jurnal Inovasi Penelitian dan Pembelajaran Fisika*, *3*(1), 10–17. https://doi.org/10.26418/JIPPF.V3I1.44544

Alfarisa, F., & Purnama, D. N. (2019). Analysis of items on the final semester exam in high school economics using the Rasch model. *Jurnal Pendidikan Ekonomi Undiksha*, *11*(2), 366–374. https://doi.org/10.23887/JJPE.V11I2.20878

Alpusari, M. (2015). Analisis Butir soal konsep dasar ipa 1 melalui penggunaan program komputer anates versi 4.0 for windows [Item analysis on the basic concepts of Science 1 through the use of the computer program Anates version 4.0 for Windows]. *Primary: Jurnal Pendidikan Guru Sekolah Dasar*, *3*(2), 106–115. https://doi.org/10.33578/jpfkip.v3i2.2501

Anggraini, D. & Suyata, P. (2014). Karakteristik soal uasbn mata pelajaran bahasa indonesia di daerah istimewa yogyakarta pada tahun pelajaran 2008/2009. *Jurnal Prima Edukasia*, 2(1), 57-65. http://dx.doi.org/10.21831/jpe.v2i1.2644

Arikunto, S. (2006). *Fundamentals of educational evaluation (revised edition)*. Yogyakarta: PT Bumi Aksara.

Astuti, B., Purwanta, E., Ayriza, Y., Bhakti, C. P., Lestari, R., & Herwin, H. (2022). School connectedness instrument's testing with the Rasch model for high school students during the COVID-19 pandemic. *Cypriot Journal of Educational Sciences*, *17*(2), 410–421. https://doi.org/10.18844/cjes.v17i2.6828

Azizah, & Wahyuningsih, S. (2020). The use of the Rasch model for the analysis of test instruments in Actuarial Mathematics course. *Jurnal Pendidikan Matematika (JUPITEK)*, *3*(1), 45–50. https://doi.org/10.30598/JUPITEKVOL3ISS1PP45-50

Banawi, A., Syafit, M., Basta, I., Ridwan, M., & Sulaeman. (2022). Evaluating educational concepts mastery of prospective teachers in Ambon: an-ex-post-facto study from comprehensive examination result. *Jurnal Pendidikan Progresif*, *12*(2), 621–634. https://doi.org/10.23960/JPP.V12.I2.202217

Banawi. (2021). *Semester lesson plan for the Basic Science Concept course*. Program Studi Pendidikan

Guru Madrasah Ibtidaiyah, Fakultas Ilmu Tarbiyah dan Keguruan, Institut Agama Islam Negeri Ambon.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 1-25. https://doi.org/10.1080/0969594X.2018.1441807

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Amsterdam, Netherlands: Springer. https://doi.org/10.1007/978-94-007-6857-4

Dewi, S. Z., & Ibrahim, H. T. (2019). The importance of understanding concepts to overcome misconceptions in Science Learning materials in Elementary Schools. *Jurnal Pendidikan UNIGA*, *13*(1), 130–136. https://doi.org/10.52434/JP.V13I1.823

Dwinata, A. (2019). Analysis of problem solving abilities using Rasch modeling on permutations and combinations materials. *Prisma, Prosiding Seminar Nasional Matematika*, *2*, 124–131. https://journal.unnes.ac.id/sju/index.php/prisma/article/view/28899

Englehard, G. (2013). *Invariant measurement, using rasch models in the social, behavioral and health sciences*. Routledge.

Erfan, M., Archi Maulyda, M., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Quality analysis of the ability to distinguish series and parallel circuits through classical test theory and the Rasch model. *Indonesian Journal of Educational Research and Review*, *3*(1), 11–19. https://doi.org/10.23887/IJERR.V3I1.24080

Fatmawati, L. (2016). Peningkatan keaktifan dan pemahaman konsep IPS melalui model learning cycle 5E. *Jurnal Prima Edukasia*, 4(2), 148-162. doi:http://dx.doi.org/10.21831/jpe.v4i2.9488

Fisher Jr., W. P. (2007). *Rating scale instrument quality criteria*. Rasch Measurement Transaction, 21, 1095. http://www.rasch.org/rmt/rmt211a.htm

Ghozali, I. (2006). *Multivariate analysis application with SPSS program*. Semarang: Badan Penerbit UNDIP.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. SAGE Publications Inc.

Hartono, W., Hadi, S., Rosnawati, R., & Retnawati, H. (2022). Test the fit of the logistic parameter model for diagnosing basic mathematical abilities. *JNPM (Jurnal Nasional Pendidikan Matematika)*, *6*(1), 125–144. https://doi.org/10.33603/JNPM.V6I1.5899

Herwin, H., Tenriawaru, A., & Fane, A. (2019). Math elementary school exam analysis based on the rasch model. *Jurnal Prima Edukasia*, *7*(2), 106-113. doi:https://doi.org/10.21831/jpe.v7i2.24450

IAIN Ambon. (2018). *Quality procedures: standard operating procedures*. Ambon: Lembaga Penjaminan Mutu Institut Agama Islam Negeri Ambon.

Ilfiandra., Nadhirah, N. A., Suryana, D., & binti Ahmad, A. (2022). Development and validation peaceful classroom scale: Rasch model analysis. *International Journal of Instruction*, 15(4), 497-514. https://doi.org/10.29333/iji.2022.15427a

Jackson, T. R., Draugalis, J. R., Slack, M. K., Zachry, W. M., & Agostino, J. D. (2002). Validation of authentic performance assessment: A process suited for rasch modeling. *American Journal of Pharmaceutical Education*, 66, 233–243.

Jusuf, R. (2015). Analysis of the ability of first semester Elementary Teacher Education students in solving questions in the basic science concept course on the subject of respiratory system using the demonstration method. *Primary: Jurnal Pendidikan Guru Sekolah Dasar*, *4*(1), 46–53. https://doi.org/10.33578/jpfkip.v4i1.2722

Jusuf, R., Sopandi, W., Ratnawulan, A., & Sa'ud, U. S. (2018). Characteristic profile of science national exam questions based on revised bloom taxonomy. *Visipena*, *9*(1), 115–127. https://doi.org/10.46244/VISIPENA.V9I1.436

Kasmo, S. (2013, August 27). *Benefits of assessment results (learning outcomes)*. Kompasiana. https://www.kompasiana.com/sukasmo/552bac716ea834bf5b8b4568/manfaat-hasil-penilaian-hasil-belajar

Khong, W. S., & Lim, H. L. (2019). The use of the Rasch model to analyze misconceptions for the topic of integer numbers. *Asia Pacific Journal of Educators and Education*, *34*, 105–128. https://doi.org/10.21315/apjee2019.34.6

Kumaidi. (1999). Application of rasch theory to vertical scaling of quarterly tests. *Jurnal Ilmu Pendidikan*, *6*(2), 89–102. https://doi.org/10.17977/JIP.V6I2.705

Lestari, W. T., Saputro, S., Masykuri, M., Hastuti, B., Ulfa, M., Mulyani, S., & Yamtinah, S. (2020).

Item analysis of teachnological pedagogical content knowledge (TPACK) in pre-service chemistry teachers using the Rasch Model application. *Journal of Physics: Conference Series*, *1511*(1), 012043. https://doi.org/10.1088/1742-6596/1511/1/012043

Ma'rifah, D. R. (2017). Diagnosis of student learning difficulties in student development courses. *Jurnal Pendidikan Biologi Indonesia*, *3*(1), 88–94. http://download.garuda.kemdikbud.go.id/article.php?article=640463&val=7984&title=Diagnosis of student learning difficulties on learner development studies

MacCann, R. G. & Stanley, G. (2006). The use of rasch modeling to improve standard setting. *Practical Assessment, Research, and Evaluation*: 11(2), 1-17. https://doi.org/10.7275/sbnk-w656

Mardapi, D. (2008). *Techniques for preparing test and nontes instruments*. Solok: Mitra Cendikia Press.

Meyer, J. P., & Hailey, E. (2012). A study of Rasch, partial credit, and rating scale model parameter recovery in WINSTEPS and jMetrik. *Journal of Applied Measurement*, *13*(3), 248–258. https://psycnet.apa.org/record/2013-00411-003

Muslihin, H. Y., Suryana, D., Ahman., Suherman, U., & Dahlan, T. H. (2022). Analysis of the reliability and validity of the self-determination questionnaire using rasch model. *International Journal of Instruction*, 15(2), 207-222. https://doi.org/10.29333/iji.2022.15212a

Natsir, N. A. (2010). The ability level of Tarbiyah Faculty Lecturers in designing instructions as participants in higher education instructional design workshops in IAIN Ambon. *Fikratuna-Jurnal Penelitian Sosial Keagamaan*, *2*(2), 21–40.

Nur, L., Nurani, L. A., Suryana, D., & Ahmad, A. (2020). Rasch model application on character development instrument for elementary school students. *International Journal of Learning, Teaching and Educational Research*, 19(3), 437-459. https://doi.org/10.26803/ijlter.19.3.24

Nurharyanto, D., & Retnawati, H. (2020). The difficulties of the elementary school students in solving the mathematical narrative test items. *Jurnal Prima Edukasia*, 8(1), 29-39. doi:https://doi.org/10.21831/jpe.v8i1.29969

Prasetyo, T. (2017). Pengembangan perangkat penilaian hasil belajar dalam pembelajaran tematik integratif kelas V SD. *Jurnal Prima Edukasia,* 5(1), 102-111. doi:http://dx.doi.org/10.21831/jpe.v5i1.7528

Purba, S. E. D. (2018). Analysis of the Rasch model of achievement test instruments on subjects of basic and electricity measurements. *Wiyata Dharma: Jurnal Penelitian Dan Evaluasi Pendidikan*, *6*(2), 142–147. https://doi.org/10.30738/WD.V6I2.3393

Purwanto, M. N. (1992). *Principles and techniques of teaching evaluation*. PT Remaja Rosdakarya.

Ramadhani, R., & Fitri, Y. (2020). The validity of the epub3-based mathematics e-module using the Rasch model analysis. *Jurnal Gantang*, *5*(2), 95–111. https://doi.org/10.31629/jg.v5i2.2535

Ramdani, R., Hanurawan, F., Ramli, M., Lasan, B. B., & Afdal, A. (2020). Development and validation of indonesian academic resilience scale using rasch models. *International Journal of Instruction*, 14(1), 105-120. https://doi.org/10.29333/iji.2021.1417a

Retnawati, H. (2014). *Item response theory and its applications: for researchers, measurement and testing practitioners, graduate students*. Nuha Medika.

Rochman, S., & Hartoyo, Z. (2018). Analysis of high order thinking skills (hots) taxonomy analyses physics problems. *SPEJ (Science and Physic Education Journal)*, *1*(2), 78–88. https://doi.org/10.31539/SPEJ.V1I2.268

Safari. (2003). *Learning evaluation*. Departemen Pendidikan Nasional, Direktorat Jenderal Pendidikan Dasar dan Menengah, Direktorat Tenaga Kependidikan.

Setiawan, A., Fajaruddin, S., & Andini, D. (2019). Development an honesty and discipline assessment instrument in the integrated thematic learning at elementary school. *Jurnal Prima Edukasia*, 7(1), 9- 19. doi:https://doi.org/10.21831/jpe.v7i1.23117

Sudjana, N. (1991). *Assessment of instructional process results*. Yogyakarta: PT Remaja Rosdakarya.

Sugiyono. (2013). *Quantitative, qualitative and R&D research methods*. Bandung: Alfabeta.

Sulistiawati, Sulistyowati, R., & Lefudin. (2021). Efforts to improve conceptual understanding in the basic science concepts course in the elementary school teacher education study program (PGSD) through remedial teaching. *PEDAGOGOS: Jurnal Pendidikan*, *3*(2), 1–10. https://doi.org/10.33627/GG.V3I2.507

Sumintono, B., & Widhiarso, W. (2015). *Rasch modeling application on educational assessment*. Trim Komunikata.

Surya, Y. F., & Marta, R. (2017). Analysis of learning difficulties of PGSD STKIP Pahlawan Tuanku Tambusai students in the subject of Elementary Basic Science Concepts in the 2016/2017 academic year. *Jurnal Basicedu*, *1*(2), 1–11. https://doi.org/10.31004/BASICEDU.V1I2.164

Syadiah, A.N. & Hamdu, G. (2020). Rasch analysis for critical thinking test questions on STEM learning in elementary schools. *Premiere Educandum: Jurnal Pendidikan Dasar dan Pembelajaran*, *10*(2), 138 – 148. Doi.org/10.25273/pe.v10i2.6524.

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. https://doi.org/10.12973/iji.2018.11210a

Widhiarso, W. (2011). Application of item response theory to test the invariance of psychological measurements for survey and job selection purposes. *Psikobuana*, 104–117. https://www.e-jurnal.com/2015/09/aplikasi-teori-respon-butir-untuk.html

Winarti, P. (2021). Analysis of students' learning difficulties in online lectures on the basic concepts of physics during the Covid-19 pandemic. *Jurnal Komunikasi Pendidikan*, *5*(1). https://doi.org/10.32585/jkp.v5i1.1076

Yusron, E., & Sudiyatno, S. (2021). How is the impact of Assessment for Learning (AfL) on mathematics learning in elementary schools?. *Jurnal Prima Edukasia*, 9(1), 75-84. doi:https://doi.org/10.21831/jpe.v9i1.34865