



Rasch model: Quality of final semester assessment items for class x on biology subject

Bekti Isnaeni*, Nani Aprilia, Much Fuad Saifuddin

Biology Education, Faculty of Teacher Training and Education, Universitas Ahmad Dahlan, Indonesia

*Corresponding author: bekti1700008098@webmail.uad.ac.id

ARTICLE INFO	ABSTRACT
<p>Article history Received: 10 August 2021 Revised: 25 July 2022 Accepted: 29 September 2022</p> <p>Keywords: Final semester Assessment MCML Multiple-Choice Card Review Rasch Model Winsteps</p>	<p>The Final The quality of the question item is important to know to produce more accurate measurements. Final Semester Assessment questions have not been analyzed, and many students still have not passed the Minimum Criterion of Mastery Learning (MCML). This research aims to determine the quality of the question items in the Final Semester Assessment of State High School in the Gondokusuman District. This research uses a quantitative descriptive approach. The population and sample in this study were the answer sheets of all students of class X MIPA at 6th and 9th State Senior High School Yogyakarta in the 2020/2021 school year, which amounted to 462 sheets. Sampling techniques used saturated samples. The data collection procedures were observation and documentation; The analysis was quantitative descriptive. Research results showed the results as 1) there is a difference in the level of question validity of the two schools; 2) The reliability of the questions whether by 6th and 9th State Senior High School Yogyakarta had excellent value; 3) The difficulty level of questions at SMA N 6 Yogyakarta is higher than SMA N 9 Yogyakarta; 4) Differentiating level of the questions by 6th State Senior High School Yogyakarta was seven groups of questions and the questions by 9th State Senior High School Yogyakarta was 11 groups of questions; 5) The effectiveness of distracting questions at 6th and 9th State Senior High School Yogyakarta is not practical; 6) The suitability of the questions with the material and construction aspects of the questions at SMA N 9 Yogyakarta is better than SMA N 6 Yogyakarta.</p>

© 2022 Universitas Negeri Jakarta. This is an open-access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0>)

Isnaeni, B., Aprilia, N., & Saifuddin, M. F. (2022). Rasch model: Quality of final semester assessment items for class x on biology subject. *Biosfer: Jurnal Pendidikan Biologi*, 15(2), 355-367. <https://doi.org/10.21009/biosferjpb.23326>

INTRODUCTION

Evaluation is one of the stages that must be taken in a learning activity. Rahmadhani (2014) says to the role of the teacher as an evaluator is very important to conduct a good and objective evaluation, besides that in the pedagogic field, the teacher must also be able to compile quality questions. An evaluation tool or test must have good quality, so as not to have an impact on measuring the ability of students. A good test that can be used is that it must contain reliable, discriminating power, and a good level of difficulty (Arvianto, 2016).

Final Semester Assessment is a very important thing to do. The purpose of implementing the Final Semester Assessment is as a form of evaluation to measure the achievement of student learning competencies that have been taught by the teacher for one semester. Susanto et al., (2015), said that the final semester score is an illustration of the mastery of competencies learned by students in the learning process at school for one semester, so good quality questions are needed.

Based on interviews with one biology teacher at State Senior High Schools in Gondokusuman, questions of the final exams were designed by Biology teachers and have never been analyzed, so the quality is not yet known. Hasanah et al. (2016), the final exam questions as a measuring tool need to be analyzed before being tested on students. Based on the description of the problem, the Biology teacher in Gondokusuman District has not taken steps to develop questions according to standards. Steps are needed to develop questions that are following standards to get quality questions. The quality of the questions can be known if they have analyzed the questions (Rahmadhani, 2014)

Based on the review document result of the odd final exam for Semester 2020/2021, many students still have not passed the Minimum Criterion of Mastery Learning (MCML). The 10% of 6th State Senior High School Yogyakarta and 83.88% of 9th State Senior High School Yogyakarta have not passed MCML. Kunandar (2014) if most of the student score below the MCML, it can be caused by questions that are arranged difficult and do not refer to the substance. It could also be that the learning carried out by the teacher cannot be understood by the students. Meanwhile, suppose almost all students get very high scores. In that case, there are several possibilities, such as questions that are arranged too easily, questions that do not follow the rules of making good questions, and the implementation is very loose, allowing students to cooperate or cheat.

One way to determinate students' abilities and results are to do an item analysis. Lubis and Prastowo (2017) said that the analysis of the quality of the items was critical to measure the achievement of student competence. The item analysis consists of two, namely qualitative analysis and quantitative analysis. Febriani (2016) said that item analysis is carried out to test each item's feasibility level based on the difficulty level and distinguishing power of the question because not all items are considered suitable for use. Item revision is based not only on the difficulty level index and differentiating power of the questions but also on the effectiveness of the distractors for each item.

In this study, researchers used the Rasch model analysis with the help of Winsteps software version 4.5.2 to analyze quantitative data, while qualitative data used a test card. The measurement results using the Rasch model can be calibrated; besides, it is not deterministic, so it can identify the object being measured carefully (Sumintono & Widhiarso, 2013). Rasch model can be used to simultaneously analyze the validity, reliability, suitability of persons and items. The Rasch model has advantages over other methods (Tenant et al., 2004; American Educational Research Association, 2014), especially Classical Test Theory (CTT), which can provide a linear scale with the same interval, is able to predict missing data so that the analysis results will be more accurate, able to produce standard error measurement values on the instrument so that it can increase the accuracy of calculations, be able to detect model inaccuracies, and can produce replicable measurements (Stolt et al., 2022). This study aims to determine the quality of the items in the Final Semester Assessment of State Senior High School in the Gondokusuman District which includes validity, reliability, difficulty level, discriminating power, and distractor effectiveness.

METHODS

Research Design

This research is included in the quantitative descriptive research, using data from students answer sheets of the final semester assessment.. The research was conducted at 'Universitas Yogyakarta Dahlan which is located at Jalan Ring road selatan, Tamanan, Banguntapan, Bantul, Special Region of

Yogyakarta. The time of research was carried out in April-May 2021 for the 2020/2021 academic year. The student answer sheets of the final semester assessment taken from all students of class X MIPA at state senior high school Yogyakarta in Gondokusuman sub-district for the academic 2020/2021 have not yet been analyzed; a total of the student answer sheets 462. Data collection techniques used are observation and documentation. I. The data analysis technique used the Rasch model with software Winstep version 4.5.2. The data of the final semester assessment be analyzed for validity, reliability, and suitability of person and items simultaneously. The data analysis technique used is descriptive qualitative and quantitative analysis.

Population and Samples

In the sub-district Gondokusuman, there are 3 State Senior High SYogyakarta, namely state senior high school 3rd of Yogyakarta, State senior high school 6th of Yogyakarta, and state senior high school 9th of Yogyakarta. However, the student's answer sheet for the Final Semester Assessment of class X MIPA at 3rdYogyakartasenior High School Yogyakarta had been analyzed. The sampling technique used was saturated sampling, namely State senior high school of Yogyakarta in sub-district Gondokusuman students' answer sheets for the final semester assessment not yet analyzed. Data collection techniques used observation and documentation. The data analysis technique used descriptive qualitative and quantitative analysis.

Table 1.

Number of Students in Class X MIPA at 6th and 9th State Senior High School Yogyakarta.

No	School name	Class	The number of students	Σ
1.	6th State Senior High School Yogyakarta	X MIPA 1	36	248 Students
		X MIPA 2	36	
		X MIPA 3	35	
		X MIPA 4	36	
		X MIPA 5	36	
		X MIPA 6	35	
		X MIPA 7	34	
2.	9th State Senior High School Yogyakarta	X MIPA 1	36	214 Students
		X MIPA 2	36	
		X MIPA 3	36	
		X MIPA 4	34	
		X MIPA 5	36	
		X MIPA 6	36	

Instrument

The instruments used in this study were multiple-choice test cards, questions, and answer sheets. The data obtained were in the form of grids, questions, student answer sheets, and answer keys for the Final Semester Assessment of Biology for class X MIPA at 6th and 9th State Senior High School Yogyakarta. Two lecturers of the Biology Education study program validated the question study card. The multiple-choice card review can be shown in [Table 2](#).

Table 2.

The Multiple-Choice Card Review.

Question number:	
No.	Review Criteria
A.	MATERIAL
1.	Question according to the indicator
2.	Deterrence works
3.	Have the correct/most correct answer
B.	CONSTRUCTION
4.	The subject matter is formulated clearly and firmly
5.	The formulation of the question and the formulation of the answer are only required statements
6.	The subject matter does not give a clue about the correct answer
7.	The subject matter does not contain double negative statements
8.	Homogeneous and logical answer choices in terms of material

Question number:	
No.	Review Criteria
9.	The length of the answer formulation is relatively the same
10.	The answer choices do not contain the statement "All of the answer choices above are wrong" or "All of the answer choices above are correct"
11.	The choice of numbers in the form of numbers or time is arranged based on the order of the size of the value of the number of chronological time
12.	Pictures, graphs, tables, diagrams, and the like are clear and functional
13.	Items do not depend on the previous question
C. LANGUAGE	
14.	The question uses language that is following the rules of Indonesian
15.	The language used is communicative
16.	Do not use the local language
17.	Answer choices do not repeat words/phrases that are not have a unified meaning

(Source: Kunandar, 2014)

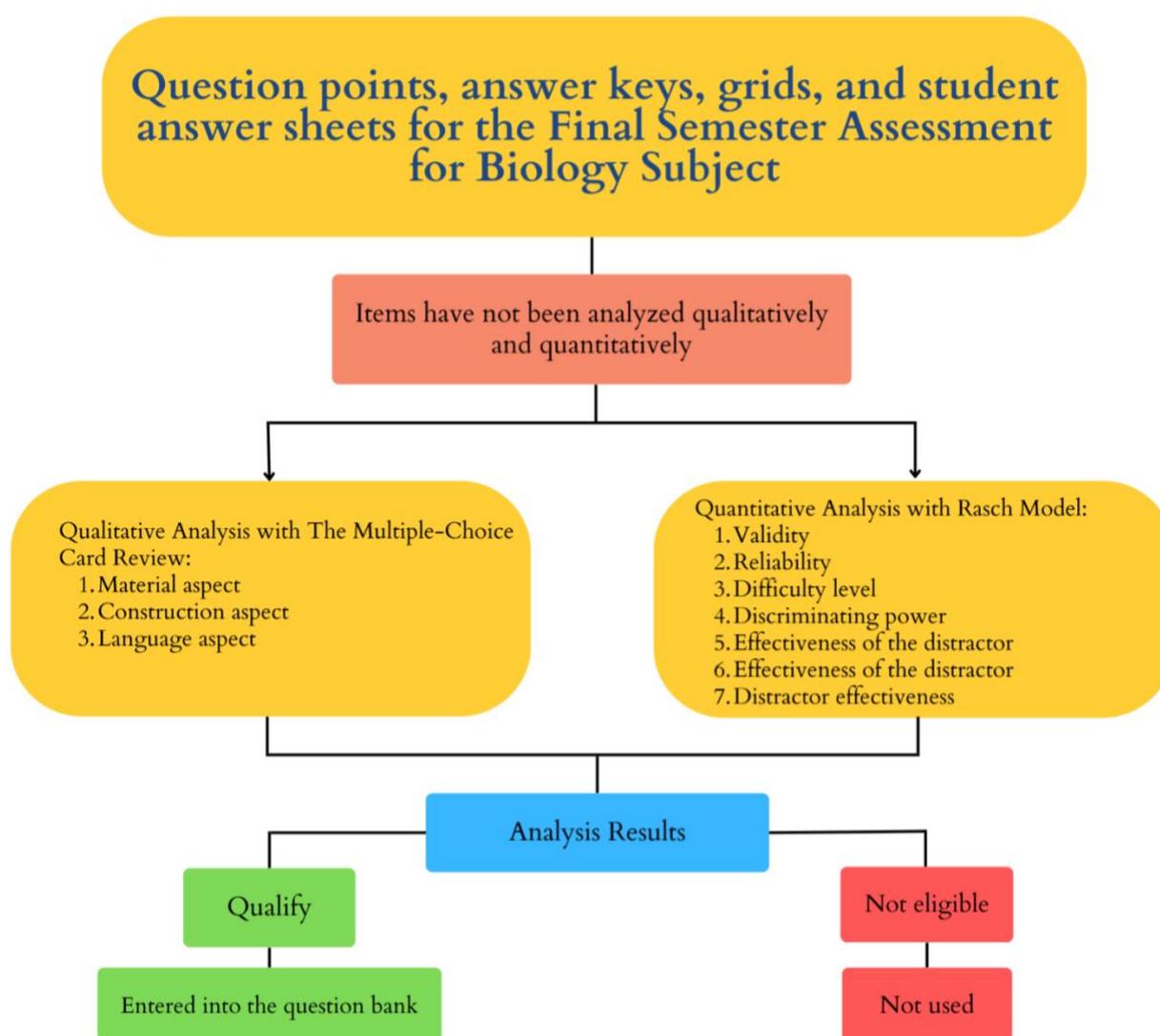


Figure 1. Research Procedure

Procedure

In the first stage, the researcher asked permission to conduct research at 6th and 9th State Senior High School Yogyakarta. In the second stage, the researcher interviewed biology teachers at 6th and 9th State Senior High School Yogyakarta to determine whether the questions Final Semester Assessment for class X Biology subject for the 2020/2021 academic year had been analyzed or not. In the third stage,

the researcher collects data using grids, questions, answer keys, and student answer sheets. In the fourth stage, the researcher conducted a qualitative and quantitative analysis. Qualitative analysis uses a question card review. The analyzed aspects are based on material, construction, and language. Quantitative analysis using Winsteps software version 4.5.2. *Software* Winsteps is a tool in the Rasch Model to analyze scores generated from test instruments. Winsteps software can determine Outfit MNSQ, Outfit ZSTD, Point Measure Correlation, Item Reliability, and Alpha Cronbach. The MNSQ outfit describes the suitability of th' data with the model used. Cronbach's Alpha value describes the reliability of the items (Azizah & Wahyuningsih, 2020). The aspects seen are validity, reliability, level of difficulty, discriminating power, and distractor effectiveness. If the analysis result meet the requirements, the item can be entered into the question bank, while the item if not used if it does not meet the requirements. In more detail, the research procedure is described in [Figure 1](#).

Data Analysis Techniques

The data analysis technique in this research is descriptive qualitative and quantitative. The data analyzed is the Final Semester Assessment of class X for the academic year 2020/2021. The total questions at 6th State Senior High School Yogyakarta are 50 questions involving 248 students and 40 questions with 214 students at 9th State Senior High School Yogyakarta. The qualitative descriptive analysis uses a question card covering material, construction, and language aspects, while the quantitative descriptive analysis uses Winsteps software version 4.5.2. The aspects are validity, reliability, level of difficulty, differentiating power, and distractor effectiveness. Item fit order (level of item suitability) indicates validity, Cronbach's alpha value indicates reliability, the item size indicates the difficulty level of the item, and the separation value indicates the distinguishing power. Furthermore, the amount of data in the category/choice/frequency nuisance table: in the order of size indicates the effectiveness of the distractor.

RESULTS AND DISCUSSION

1. Quantitative Analysis

a. Validity

Item fit order (level of item suitability) indicates validity. The results of the item validation analysis can be seen in [Table 3](#).

Table 3.

The Results of the Validity of the Final Semester Assessment Items at 6th and 9th State Senior High School Yogyakarta

School	Question Number	Amount	Percentage	Information
6th State Senior High School Yogyakarta	38, 36, 46, 44, 24, 43, 18, 41, 50, 23, 30, 26, 29, 31, 35, 33, 40, and 49	18	36%	Valid
	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 22, 28, 32, 34, 47, 25, 45, 19, 48, 37, 1, 39, 42, and 27	32	64%	Invalid
9th State Senior High School Yogyakarta	7, 18, 15, 35, 14, 10, 22, 24, 28, 32, 34, 1, 4, 5, 6, 31, 19, 36, 37, 8, 11, 12, 21, 13, 16, 20, 26, 29, 23, 9, 2, and 33	32	80%	Valid
	30, 38, 39, 40, 3, 17, 27, and 25	8	20%	Invalid

Based on [Table 3](#), the validity quality questions of the final assessment semester in the ninth state senior high school of Yogyakarta are better than in the sixth state senior high school of Yogyakarta. The item state valid can view from the amount of Infit and Outfit values that are maximum measured. The question item is declared fit if the MNSQ Outfit value is between 0.5 to 1.5; Outfit ZSTD values are between -2.0 to +2.0; and the correlation point measure value is between 0.4 to 0.85 (Sumintono & Widhiarso, 2015). Widyaningsih and Yusuf (2018) the problem of misfit means having indications of misconceptions in understanding and having a problem working questions.

Rahmani et al. (2015) said the item questions are valid if they can measure the expected competence, while the item questions are invalid if the cannot measure the expected competence. So it

can be known that in the final semester assessment of Biology subjects at 6th State Senior High School Yogyakarta, as many as 18 questions (36%) are valid. Whereas at 9th State Senior High School, Yogyakarta, as many as 32 questions (80%) are valid.

Based on the description, there is a difference between the validity of the Final Semester Assessment at 9th State Senior High School Yogyakarta is very high than at 6th State Senior High School Yogyakarta. Factors that cause the difference are the number of students and the number of questions. The number of students and item questions in 6th State Senior High School Yogyakarta is more than in 9th State Senior High School Yogyakarta. So, the Final Semester Assessment of Biology subjects at 6th State Senior High School Yogyakarta has poor validity.

In addition to the validity aspect of the item, it needs to be analyzed unidimensionality. According to Misbach and Sumintono (2014), unidimensionality is important to know whether the instrument developed can measure what should be measured. Based on results, the raw variance measurement results are 32.9% and 43.1%. This result indicates that the minimum unidimensionality requirement of 20% has been met. In addition, obtained variances that instruments cannot explained show independence in instruments of good value because it is below 7% (Misbach & Sumintono, 2014). Thus, the instrument developed is valid enough to measure student's abilities (Novinda et al., 2019).

b. Reliability

Cronbach's alpha value indicates reliability. The results of the reliability analysis of the problem can be seen in Table 4.

Table 4.

Summary of the Final Semester Assessment Items at 6th and 9th State Senior High School Yogyakarta.

Information		Score	
		6th State Senior High School Yogyakarta	9th State Senior High School Yogyakarta
		Logit	Person
	Items	0.00	0.00
Reliability	Person reliability	0.21	0.64
	Item reliability	0.96	0.98
	Alpha Cronbach	0.47	0.62
MNSQ Outfits	Person	-	-
	Items	0.90	1.17
ZSTD Outfits	Person	-	-
	Items	-0.06	0.19

Bulqis (2019) states reliability means reliable and trustworthy because by doing reliability analysis, we can find how much consistency or determination to measure problems so that problems can be tested in school and equivalent. Based on the results of data analysis using Winsteps software, in Table 4, the person measure value is 2.70 and 0.57 logit. This grade shows all students' average grades in the problem items. Cronbach's alpha value measures reliability, indicating the interaction between the person and the overall problem items of low and destructive values of 0.47 and 0.62. Cronbach's alpha value describes data that does not vary, but this value does not affect validity.

The person reliability values in Table 4 are 0.21 and 0.64, whereas reliability items are 0.96 and 0.98. Hasanah et al. (2016), said that the reliability value getting closer to the number 1 than the reliability value of the test or test is higher / better. It can be known that the consistency of answers from students is weak, but the quality of the problem items in the instrument is exceptional. According to Wahyudi et al. (2020), finding out whether the item developed quality or not can be seen from the value of reliable items. This result shows that the quality of items for the Final Semester Assessment of Biology subjects in 6th and 9th State Senior High School Yogyakarta is outstanding because it falls above 0.94 (Sumintono & Widhiarso, 2015).

The value can also be reliable in the Rasch model of person separation and separation items. The greater the person separation value, the better the test used because it can reach respondents' ability. High values of separation items also show the better the measurements taken (Sumintono & Widhiarso, 2015). So, it can be known that the reliability of the Final Semester Assessment of Biology subjects in 6th and 9th State Senior High School Yogyakarta is excellent because it is above 0.94.

c. Level of difficulty

The item size indicates the difficulty level of the item. The results of the problem item difficulty analysis can be seen in [Table 5](#).

Table 5.

The Results of the Analysis of the Difficulty Level of the Final Semester Assessment Items at 6th and 9th State Senior High School Yogyakarta.

School	Question Number	Amount	Percentage	Information
6th State Senior High School Yogyakarta	29 and 33	2	11,11%	Very difficult
	36, 48, 18, 43, 23, 46, 24, 30, and 50	9	50%	Difficult
	40, 41, 31, and 38	4	22,22%	Easy
	35, 38 and 44	3	16,66%	Very easy
9th State Senior High School Yogyakarta	36, 14, 15, 29, and 32	5	15,62%	Very difficult
	11, 1, 28, 6, 33, 8, 19, 5, and 37	9	28,12%	Difficult
	26, 34, 22, 35, 23, 12, 2, 10, 31, 4, 13, 9, and 16	13	40,62%	Easy
	21, 7, 20, 24, and 18	5	15,62%	Very easy

Sari and Herawati (2014) stated that the level of difficult questions is to measure the ease and difficulty of the problem to be tested. Analysis of the level of difficulty of the questions is very important because it is used to calibrate the questions in determining the questions in easy, medium, and difficult criteria so that it can consider the proportion of each criterion in the prepared question sheet. The difficulty level in Rasch model analysis can be known by looking at the output of the item measure table. The problem's difficulty level based on the Rasch model is determined mainly by the student's response/answer to the problem; this distinguishes it from conventional analysis (Hamdu et al., 2020).

According to Sabekti and Khoirunnisa (2018), the difficulty level of problem items can be classified by comparing the measured value of each problem item with the value of the S.D measure. Misbach and Sumintono (2014), in their research, said that if obtained the average logit item is not 0.0, then overall, the instrument is not good. After screening the problem of misfits and outliers obtained, the results of the analysis of the level of difficulty in the Final Semester Assessment Gasal Biology subjects class X at 6th State Senior High School Yogyakarta can be categorized into four categories, namely, as many as 2 points of questions with a percentage of 11.11% including very difficult problems, 9 points of questions with a percentage of 50% including difficult problems, 4 points of questions with a percentage of 22.22% including easy questions, and 3 points of questions with a percentage of 16.66% including very easy problems. The results of the analysis of the Final Semester Assessment of Biology class X subjects at 9th State Senior High School Yogyakarta are categorized into four categories, namely as many as five questions with a percentage of 15.62% including very difficult problems, 9 questions with a percentage of 28.18% including difficult problems, 13 questions with a percentage of 40.62% including easy questions, and 5 points of questions with a percentage of 15.62% including very easy problems.

Syadiah and Hamdu (2020) said that the high difficulty level of questions could be seen from the highest logit values. This condition corresponds to the total score column representing how many respondents answered correctly on the tested question (Widyaningsih & Yusuf, 2018). According to Irmalasari et al., (2016), this follows the theory that if the difficulty level is lower than the student's ability, then the questions is relatively easy. Conversely, the questions is relatively tricky if the difficulty level is higher than the student's ability. If student's level of difficulty and ability is balanced, then the questions is classified as moderate. According to Erfan et al. (2020), a good quality questions is a question that is not too difficult and not too easy. Questions that have a low difficulty level with logit values below -1 must be revised again (Ibnu et al., 2019).

d. Distinguishing power

The separation value indicates the distinguishing power. The results of the analysis of the distinguishing power can be seen in [Table 6](#).

Table 6.

The Results of the Distinguishing Power of Final Semester Assessment Items at 6th and 9th State Senior High School Yogyakarta.

Variable	Value separation	
	6th State Senior High School Yogyakarta	9th State Senior High School Yogyakarta
Person	0.52	1.34
Item	4.66	7.65

According to Sumintono and Widhiarso (2015), grouping the distinguishing power of questions more accurately is called strata separation. Kunandar (2014) stated that distinguish of a question is the ability of a question to distinguish whether student's have mastered the material or not. Rasch model does not contain parameters of discrimination, all points of the questions are determined to have equality in the power of discrimination. The item's difficulty level is the only item parameter focused on the Rasch model (Sumintono & Widhiarso, 2015). Alfarisa and Purnama (2019) said that the Rasch model with (1PL) characteristics of visible question items is the difficulty level of the grain, while the other power is considered constant.

Sumintono and Widhiarso (2015) say the grouping of distinguishing power of questions is more thoroughly called strata separation (H). Ibnu et al. (2019) said that the greater the value of separation, the quality of the question item instrument used is very good because it can identify the group of question points with the group of respondents. Susdelina et al., (2018) stated that the analysis of Rasch's model differs from classical test theory to distinguish student's high and low ability using analysis at the level of individual ability. In addition, it can be seen by identifying the group of respondents based on the respondent separation index. The results of the analysis of the distinguishing power of the question using the Rasch model can be seen in the output of the summary statistics table in the separation column.

Based on the analysis of the distinguishing power of the Final Semester Assessment of Biology subjects at 6th State Senior High School Yogyakarta obtained seven questions based on the index of separation of points of $H = 6.54$. The results of the distinguishing power analysis of the Final Assessment of Semester of Biology subjects at 9th State Senior High School Yogyakarta there are eleven groups of questions based on the index of separation of points of questions $H = 10.53$.

e. Distractor effectiveness

The amount of data in the category/choice/frequency nuisance table: the order of size indicates the effectiveness of the distractor. The effectiveness of distractors on the final assessment semester of Biology subjects in class X at 6th and 9th State Senior High School Yogyakarta can be seen in Tables 7 and 8.

Table 7.

The Results of the Analysis of the Distractor Effectiveness of Final Semester Assessment Items at 6th State Senior High School Yogyakarta.

Question number	Answer key	Count					Distractor Description
		A	B	C	D	E	
33	A	68	87	60	16	17	4 working
49	B	13	208	5	12	10	2 works
18	C	0	0	212	5	31	1 works
23	C	0	25	222	0	1	1 works
24	A	220	16	8	0	4	1 works
29	C	83	10	144	1	9	1 works
30	B	14	222	10	2	0	1 works
36	B	10	194	31	10	3	1 works
40	B	0	227	18	2	1	1 works
43	D	30	0	0	216	2	1 works
46	E	2	17	8	4	217	1 works
50	A	222	12	6	5	3	1 works
31	C	1	10	237	0	0	does not work
35	C	3	1	244	0	0	does not work
38	B	3	241	0	1	3	does not work

Question number	Answer key	Count					Distractor Description
		A	B	C	D	E	
41	A	231	1	4	7	5	does not work
44	A	243	2	2	1	0	does not work

Table 8.

The Results of the Analysis of the Distractor Effectiveness of Final Semester Assessment items at 9th State Senior High School Yogyakarta.

Question number	Answer key	Count					Distractor Description
		A	B	C	D	E	
14	C	32	140	4	24	13	4 working
19	D	18	33	27	111	24	4 working
28	B	59	63	41	21	29	4 working
29	E	53	30	28	64	37	4 working
5	E	22	42	4	18	127	3 works
6	D	8	40	81	73	11	3 works
15	C	97	42	26	3	45	3 works
23	C	21	11	161	10	11	3 works
36	D	11	14	180	10	6	3 works
2	D	15	10	4	165	19	2 works
8	B	7	110	74	16	6	2 works
10	B	17	166	23	2	3	2 works
11	E	39	116	3	8	47	2 works
12	C	29	12	164	4	3	2 works
13	B	17	179	4	12	1	2 works
26	B	7	139	4	46	18	2 works
34	A	146	6	37	19	6	2 works
37	C	9	7	134	51	12	2 works
1	A	52	0	3	8	150	1 works
4	C	0	7	176	5	23	1 works
9	B	1	188	23	0	1	1 works
16	D	11	3	3	195	2	1 works
22	D	4	42	2	157	9	1 works
27	A	156	51	5	1	1	1 works
31	B	24	168	7	4	10	1 works
33	A	101	4	108	1	0	1 works
35	D	2	41	4	3	8	1 works
7	D	0	1	5	203	4	does not work
18	A	211	0	0	2	0	does not work
20	C	2	1	204	3	4	does not work
21	D	0	2	6	200	5	does not work
24	B	4	206	0	2	2	does not work

In interpreting the effectiveness of distractors (Tables 7 and 8), used the criteria from Oktanin and Sukirno (2015). The effectiveness of the distractor can be seen in the output table of item category/option/distractor frequencies: measure order. Tables 7 and 8 shows that most student can work correctly on item questions than they can not be worked. Case and Donahue (2008) said that a distractor that works means it can produce more difficult item questions. The distractor also reduce random guessing of answers to improve the performance of the questions. Thus, it can be known that the distractor/cheater on the final semester assessment of biology subjects class X at 6th and 9th State Senior High School Yogyakarta does not function effectively. This condition can be seen from the number of distractors that work correctly.

2. Qualitative Analysis

Qualitative analysis uses a question card that covers material, construction, and language aspects. The results of the qualitative analysis can be shown in Table 9.

Table 9.

The Results of the Qualitative Analysis of the Items for the Final Semester Assessment at 6th and 9th State Senior High School Yogyakarta.

Indicator	Question Numbers that do Not Match the Indicators	
	6th State Senior High School Yogyakarta	9th State Senior High School Yogyakarta
1. Question according to the indicator (Material Aspect)	21, 22, 23, 25, 26, 30, 31, 37, 39, 41, 43, 45, 48	19 and 32
2. The length of the answer formula is relatively the same (Construction Aspect)	35 and 50	5, 10, 15, 29, 34 and 36

Sukiman (2012) said that theoretical or qualitative analysis could be done before and after the trial. Qualitative analysis is seen from 3 aspects: material, construction, and language. How to analyze it is to look at the details of the problem compiled from the fulfillment of the requirements of aspects of content (material), construction, and language.

In this study, qualitative analysis was conducted by three reviewers, namely alumni of biology education study programs, peers, and researchers, which is conducted through a panel technique. The validation process by reviewers is carried out in their respective places so that they can be objectives and between reviewers do not affect each other. Each reviewer is given review formats and assessment guidelines. Reviewers are also welcome to correct the direction on the text of the question, provide comments or suggestions, and rate each item of the questions with criteria: Good, revision, or replaced.

The results of qualitative analysis on the Final Semester Assessment of Biology subjects class X at 6th State Senior High School Yogyakarta in material aspect showed 13 question items that did not match the indicators. Whereas at 9th State Senior High School, Yogyakarta showed 2 question items. Besides, in the construction aspect, two question items (from 6th State Senior High School, Yogyakarta) and 6 question items (from 9th State Senior High School, Yogyakarta) do not match the indicator. While in the language aspect, all item questions match with the indicator.

Qualitative analysis also showed that three question numbers are not in the latticework. In compiling the question, the details of the question are adjusted to the existing latticework (Mujimin, 2010). If the question matches existing latticework, then the test results can be used to determine the actual competence of students. Ambiyar and Panyahuti (2020) said that the guidelines in preparing questions are indicators contained in the latticework. Therefore, the teacher's ability to compile the latticework of questions needs. The teacher was the most significant factor in preparing the problem (Rahmadhani, 2014). A teacher needs special abilities such as discussing question ideas, understanding the characteristics of learners, and mastery of question writing techniques so that the questions tested on learners follow standards. Lubis, and Prastowo (2017) said that a teacher must be able to arrange quality problems to know the extent to which students understand the material that has been taught.

3. Quality of Questions

The results of the questions' quality analysis are interpreted following the criteria from Oktanin and Sukirno (2015). The quality of the items for the Final Semester Assessment of 6th and 9th State Senior High School Yogyakarta can be shown in Table 10.

Table 10.

The Quality of the Items for the Final Semester Assessment of 6th and 9th State Senior High School Yogyakarta.

Category Enter the Question Bank	Question Number	
	6th State Senior High School Yogyakarta	9th State Senior High School Yogyakarta
Yes	49	2, 5, 8, 10, 12, 13, 19, 24, 34, and 37
No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 25, 27, 28, 32, 34, 37, 39, 42, 45, 47, and 48	3, 17, 25, 27, 30, 38, 39, and 40
Not Yet	18, 23, 24, 26, 29, 30, 31, 33, 35, 36, 40, 41, 43, 44, 46, and 50	1, 4, 6, 7, 9, 11, 14, 15, 16, 18, 20, 21, 22, 24, 28, 29, 31, 32, 33, 35, and 36

Based on the analysis results in [Table 10](#), item questions with the category "Yes" can be directly entered into the question bank because the quality is very good. The item questions with the category "Not Yet" need to be revised to include a question bank. However, the item questions with the category "No" this quality of item questions is terrible, so it is better not to enter them into the question bank and not use them to measure student competencies. Questions to measure student competence should have good quality to measure correctly.

CONCLUSION

The quality of the end-of-semester assessment questions in this study has very good reliability. However, some questions should be improved related to the material and question construction aspects to produce better questions to be included in the question bank as material for the assessment questions at the end of the following semester. It is expected that biology teachers conduct trials and analyses of item questions before being used as test instruments to comply with the guidelines of question development measures.

ACKNOWLEDGMENT

Researchers would like to thank all those involved in helping complete the study. Thank you to the lecturer of Biology Education Universitas Ahmad Dahlan Yogyakarta. Thank you also to the Principal and biology teacher of 6th and 9th State Senior High School Yogyakarta who have provided the opportunity to conduct this research.

REFERENCES

- Alfarisa, F., & Purnama, D. N. (2019). Analisis butir soal ulangan akhir semester mata pelajaran ekonomi SMA menggunakan Rasch model. *Jurnal Pendidikan Ekonomi*, 11(2), 366–374. <https://ejournal.undiksha.ac.id/index.php/JJPE/article/view/20878>
- Ambiyar, & Panyahuti. (2020). *Asesmen pembelajaran berbasis komputer dan android (Pertama)*. Prenada Media.
- American Educational Research Association. (2014). *The Standards for Educational and Psychological Testing*. American Educational Research Association.
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan*. PT Bumi Aksara.
- Arviyanto, I. R. (2016). Pemanfaatan program Iteman 3.0 untuk analisis butir soal lomba cerdas cermat teknologi informasi dan komunikasi tingkat SMA sederajat. *Jurnal Teknologi Informasi*, XI(33), 1–13. <https://doi.org/10.35842/jtir.v11i33.106>
- Azizah, A., & Wahyuningsih, S. (2020). Penggunaan model rasch untuk analisis instrumen tes pada mata kuliah matematika aktuaria. *JUPI TEK: Jurnal Pendidikan Matematika*, 3(1), 45–50. <https://doi.org/10.30598/jupitekvol3iss1pp45-50>
- Bulqis, A. (2019). Analisis butir soal pilihan ganda biasa ulangan tengah semester buatan guru Biologi kelas X SMA Negeri 3 Maros. *Prosiding Seminar Nasional FKIP Universitas Muslim Maros*, 1, 191–197. <https://ejournals.umma.ac.id/index.php/prosiding/article/view/377/267>
- Case, S. M., & Donahue, B. E. (2008). Developing high-quality multiple choice questions or assessment in Legal Education. *Journal of Legal Education*, 58(3), 372–387. <https://www.jstor.org/stable/42894078>
- Erfan, M., Mauliyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Analisis kualitas soal kemampuan membedakan rangkaian seri dan paralel melalui teori tes klasik dan model Rasch. *Indonesian Journal of Educational Research and Review*, 3(1), 11–19. <https://doi.org/10.23887/ijerr.v3i1.24080>
- Febriani, I. M. (2016). Analisis butir soal ujian akhir semester (UAS) bahasa jerman kelas X MIA 6 SMA negeri 1 maospati tahun ajaran 2015/2016. *Laterne E-Jurnal UNESA*, 5(2). <https://core.ac.uk/download/pdf/230655215.pdf>
- Hamdu, G., Fuadi, F. N., Yulianto, A., & Akhirani, Y. S. (2020). Items quality analysis using Rasch model to

measure elementary school students' critical thinking skill on stem learning. *JPI (Jurnal Pendidikan Indonesia)*, 9(1), 61. <https://doi.org/10.23887/jpi-undiksha.v9i1.20884>

- Hasanah, A. L., Subali, B., & Mariyam, S. (2016). Analisis item ujian akhir semester (UAS) genap mata pleajaran biologi kelas X tahun ajaran 2014/2015 di SMA yang mengimplementasikan kurikulum 2013 di kabupaten sleman. *Jurnal Pendidikan Biologi*, 5(4), 7–26. <https://journal.student.uny.ac.id/ojs/index.php/jeb/article/view/4530>
- Herrmann-Abell, C. F., & DeBoer, G. E. (2011). Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. *Chem. Educ. Res. Pract.*, 12(2), 184–192. <https://doi.org/10.1039/C1RP90023D>
- Ibnu, M., Indriyani, B., Inayatullah, H., & Guntara, Y. (2019). Aplikasi Rasch Model: Pengembangan instrumen tes untuk mengukur miskonsepsi mahasiswa. *Prosiding Seminar Nasional Pendidikan FKIP*, 2(1), 205–210. <https://jurnal.untirta.ac.id/index.php/psnp/article/view/5669>
- Irmalasari, K., Suratsih, & Wibowo, Y. (2016). Analisis butir soal ulangan akhir semester genap kelas X Biologi tahun ajaran 2015/2016. *Jurnal Pendidikan Biologi*, 5(8), 10–18. <https://journal.student.uny.ac.id/ojs/index.php/jeb/article/view/6043>
- Kunandar. (2014). *Penilaian Autentik (Penilaian Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013)*. Rajawali Pers.
- Lubis, N. S., & Prastowo, P. (2017). Analisis kualitas soal ujian mata pelajaran biologi di SMA Negeri 3 Medan tahun pembelajaran 2016/2017. *Jurnal Pelita Pendidikan*, 5(1), 149–157. <https://jurnal.unimed.ac.id/2012/index.php/pelita/article/view/8456>
- Misbach, I. H., & Sumintono, B. (2014). Pengembangan dan validasi instrumen “Persepsi siswa terhadap karakter moral guru” di Indonesia dengan Model Rasch. *Seminar Nasional Psikometri Dengan Tema “Pengembangan Instrumen Penilaian Karakter Yang Valid,”* 1–17.
- Mujimin. (2010). Kompetensi guru dalam menyusun butir soal pada mata pelajaran bahasa Jawa di sekolah dasar. *LINGUA (Jurnal Bahasa Dan Sastra)*, 6(2).
- Novinda, M. R. R., Silitonga, H. T. M., & Hamdani. (2019). Pengembangan tes pilihan ganda menggunakan model Rasch materi gerak lurus kelas X Pontianak. *Jurnal Pendidikan Dan Pembelajaran*, 8(6), 1–11. <https://jurnal.untan.ac.id/index.php/jpdpb/article/view/33452>
- Oktanin, W. S., & Sukirno. (2015). Analisis butir soal ujian akhir mata pelajaran ekonomi akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 8(1), 35–44. <https://doi.org/10.21831/jpai.v13i1.5183>
- Rahmadhani, E. K. (2014). Analisis kualitas butir soal pada bank soal biologi kelas X SMA. *Bioedu (Berkala Ilmiah Pendidikan Biologi)*, 3(1), 422–438. <https://ejournal.unesa.ac.id/index.php/bioedu/article/view/7055/7650>
- Rahmani, M., Ningsih, K., & Nurdini, A. (2015). Analisis kualitas butir soal buatan guru Biologi kelas X SMA Negeri 1 Tanah Pinoh. *Jurnal Pendidikan Dan Pembelajaran Khatulistiwa*, 4(2). <https://jurnal.untan.ac.id/index.php/jpdpb/article/view/8970/8901>
- Sabekti, A. W., & Khoirunnisa, F. (2018). Penggunaan Rasch Model untuk mengembangkan instrumen pengukuran kemampuan berpikir kritis siswa pada topik ikatan kimia. *Jurnal Zarah*, 6(2), 68–75. <https://doi.org/10.31629/zarah.v6i2.724>
- Sari, A. I. C., & Herawati, M. (2014). Aplikasi Anates versi 4 dalam menganalisis butir soal. *Faktor Jurnal Ilmiah Kependidikan*, 1(2), 203–214. <https://journal.lppmunindra.ac.id/index.php/Faktor/article/view/353>
- Sugiyono. (2019). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Alfabeta.
- Sukiman. (2012). *Pengembangan Sistem Evaluasi*. Insan Madani.
- Sukmadinata, N. S. (2015). *Metode Penelitian Pendidikan*. PT Remaja Rosdakarya.
- Sumintono, B., & Widhiarso, W. (2013). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial*. Trim

Komunikata Publishing House.

- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim Komunikata Publishing House.
- Susanto, H., Rinaldi, A., & Novalia. (2015). Analisis validitas reliabilitas tingkat kesukaran dan daya beda pada butir soal ujian akhir semester ganjil mata pelajaran matematika. *Al-Jabar: Jurnal Pendidikan Matematika*, 6(2), 203–217. <https://doi.org/10.24042/ajpm.v6i2.50>
- Susdelina, Perdana, S. A., & Febrian. (2018). Analisis kualitas instrumen pengukuran pemahaman konsep persamaan kuadrat melalui teori tes klasik dan Rasch Model. *Jurnal Kiprah*, 6(1), 41–48. <https://doi.org/10.31629/kiprah.v6i1.574>
- Stolt, M., Kottorp, A., & Suhonen, R. (2022). The use and quality of reporting of Rasch analysis in nursing research: A methodological scoping review. *International Journal of Nursing Studies*, 132, 104244. <https://doi.org/10.1016/j.ijnurstu.2022.104244>
- 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>
- Tenant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value Health*, 7, S22-S26. <https://doi.org/10.1111/j.1524-4733.2004.7s106.x>
- Wahyudi, A., Setyowati, A., & Partini, S. (2020). Analisis model Rasch pada pengembangan skala resiliensi. *Jurnal Fokus Konseling*, 6(2), 68–74. <https://doi.org/10.52657/jfk.v6i2.1157>
- Widyaningsih, S. W., & Yusuf, I. (2018). Analisis soal modul laboratorium fisika sekolah I menggunakan Rasch model. *Gravity: Jurnal Ilmiah Penelitian Dan Pembelajaran Fisika*, 4(1), 33–46. <http://dx.doi.org/10.30870/gravity.v4i1.3116>