Journal of Chemistry Education, Vol. 12 (1) 2023
Chemistry Education Study Program
Universitas Sebelas Maret

Page 62-72
ISSN 2337-9995
https://jurnal.uns.ac.id/jpkim

ANALYSIS OF FINAL SEMESTER ASSESSMENT QUESTIONS FOR CLASS XI CHEMISTRY SUBJECTS AT HIGH SCHOOL 1 WONOGIRI FOR THE 2022/2023 ACADEMIC YEAR

Clara Ria Dentawati*, Endang Susilowati, dan Sulistyo Saputro

Chemistry Education Study Program, FTTE, Universitas Sebelas Maret, Surakarta, Indonesia

*Correspondence requirements, phone: 085156806850, email: clarariad@gmail.com

ABSTRACT

This study aims to analyze the quality of the Final Semester Assessment questions for the chemistry subject of High School 1 Wonogiri using Iteman 4.5 regarding the validity, reliability, discriminatory power, difficulty level, and deceivers' effectiveness. This study used a quantitative descriptive research method. Samples are in the form of student answers. Data was obtained using documentation techniques in 1 set of questions, answer keys, and student answer responses of as many as 108. Data analysis techniques using the Iteman program version 4.5. The analysis results showed that in the aspect of validity, 76.67% of the questions were valid, and 33.34% were invalid. In the reliability aspect, an alpha reliability value of 0.82 was obtained and was in the high category. Regarding difficulty, 33.34% of the questions were obtained in the difficult category, and 66.66% of the questions were in the medium category. Regarding discriminatory power, 7% of the questions were terrible, 20% were deficient, 40% were in the medium category, and 33% were in the excellent category. In the aspect of deceptive effectiveness, 7% of the questions were categorized as sufficient, 30% of the questions were classified as good, and 63% of the questions were classified as very good.

***Keywords**: Question Item Analysis, End-Of-Semester Assessment, Iteman*

INTRODUCTION

Attempts made to develop or change intentional behavior are called education [1]. In a broad sense, education can also be defined as learning activities carried out throughout time that aim to positively impact individuals [2]. The purpose of education can be achieved by the existence of schools. In addition, it is necessary to conduct an evaluation to determine the extent of students' ability to capture and understand learning.

Learning objectives can be achieved by holding periodic and continuous evaluations. According to [3], evaluation can be interpreted as a process carried out systematically to determine quality with specific considerations. In the learning process, the duty of teachers or educators is to evaluate and develop instruments or evaluation tools adapted to measuring techniques so that the intended indicators can be achieved [4]. One of the evaluation tools is the test. According to [5], Quality instruments are essential for educators or teachers so that educators or teachers can choose good questions. Therefore, it is necessary for the teacher to analyze the question items before testing the questions on students.

Evaluation, assessment, and measurement have a very close and related relationship. For example, assessment activities cannot be separated from measurement activities. Measurement instruments are very important in determining the quality of measurement results. Good judgment is due to good instruments, while bad or misleading judgment is caused by imperfect instruments [6].

According to Permendiknas Number 23 of 2020, student learning assessment can be interpreted as processing and collecting information related to student learning outcomes

within a certain period of time. One of the assessments within a certain period of time is the daily test and the end-of-semester assessment (PAS). This assessment is usually in the form of a teacher-made test. According to [7], Tests compiled by teachers that are used to assess the learning process are called teacher-made tests. Assessments carried out by educational units, such as midterm assessments, end-of-semester assessments, or grade advancement exams, include summative assessments because they are carried out at certain times [6][8].

According to Subali (2014), question items can be analyzed in two ways: qualitative analysis and quantitative analysis [9]. According to Fatimah & Alfath (2019) Quantitative analysis has two approaches, including the classical approach and the modern approach [10]. The approach that is widely used today is the classical test theory approach (CTT). Classical test theory emphasizes a single test's visible score, usually summed up as the test taker's ability. CTT analysis is cheaper and more straightforward, but the characteristics of the resulting question items are volatile (inconsistent) depending on the test taker's ability.

In addition, the modern approach is the study of question items using Item Response Theory (IRT) or question item answer theory. One of the most famous models of contemporary test theory is Rasch. The advantages of this modeling are that it can produce accurate statistical analysis data and predict lost data. However, this modeling also has the disadvantage that mathematical equations are more difficult to understand, so they require computer assistance.

The benefits of question item analysis are to increase the validity and reliability of the questions, determine questions that do not work well, revise questions that are not relevant to the material taught, and improve the question items through analysis components such as discriminatory power, level of difficulty, and effectiveness of deceivers. It can also assist in preparing local and informal tests, such as tests prepared by educators or teachers [11].

In research conducted by Santosa (2022), shows that there is no trial on the instrument and no question analysis is carried out, so the quality of the questions tested is unknown [5]. The absence of testing the questions causes the quality of the questions to be unfulfilled as a requirement for good questions. Question items of good quality can be seen in terms of validity, reliability, discriminatory power, and difficulty level [12].

A question can be valid if it has three aspects that can be measured: skills, attitudes, and knowledge. A question can be reliable if it has reliability, which means that the question items are no different when tested on the same subject but different in time. Another aspect that is no less important in determining the questions' quality is the difficulty level and discriminatory power. Judging from the difficulty level, questions can be said to be good if they are not too difficult or not too easy. While viewed from the power of difference, a good question is a question that can distinguish students who have high and low abilities [13]. The effectiveness of the deceiver is no less critical. The question is good if students with low grades prefer deceivers compared to high score students [14]. According to Perdana (2018), testing with these aspects is absolutely carried out to obtain quality test instruments [15].

The instrument to be analyzed is the End-Semester Assessment (PAS) exam questions for the 2022/2023 academic year for the chemistry subject of High School 1 Wonogiri in the form of multiple choice questions. Based on observations, High School 1 Wonogiri is a favorite school in Wonogiri. After an interview with chemistry teachers, it was found that the PAS questions themselves were prepared by chemistry teachers from the question bank. The material has never been analyzed. If the quality of the questions is not known with certainty, it will impact misinterpretation of the test results. Of course, it will also affect the bias of information obtained from assessment tools about student abilities

RESEARCH METHOD

This study used a quantitative descriptive research method. The data collection technique uses documentation techniques in the form of 1 set of questions along with answer keys and student answer responses of as many as 108. Data analysis techniques using the help of the Iteman program version 4.5. The data analysis process describes the results of the analysis of the PAS Chemistry problem of High School 1 Wonogiri in terms of validity, reliability, difficulty level, discriminatory power, and effectiveness of deceivers.

RESULTS AND DISCUSSION

1. Validity

According to Drost (2011), validity can be interpreted as the most essential thing in the research component [16]. Ronald Jay stated that test validity aims to be a good test benchmark. More precisely, evidence-based measurements are used to determine the feasibility of a test [17]. Validity is defined as the ability of an instrument to measure what should be measured.

Table 1 Group Distribution 1-5 Question No. 1

Opt	N	Group 1	Group 2	Group 3	Group 4	Group 5
A	14	0,33	0,28	0,10	0,00	0,05
B	38	0,47	0,11	0,40	0,30	0,45
C	4	0,07	0,00	0,03	0,00	0,09
D	23	0,07	0,28	0,23	0,26	0,18
E*	**29**	**0,07**	**0,33**	**0,23**	**0,43**	**0,23**

Table 3 Option Statistic Question No. 18

Opt	N	Prop.	Rpbis	Rbis	Mean	SD
A*	**17**	**0,16**	**-0,01**	**-0,02**	**12,71**	**4,77**
B	21	0,19	-0,17	-0,24	10,00	4,81
C	20	0,19	0,06	0,09	12,55	5,76
D	5	0,05	0,06	0,13	13,40	3,78
E	45	0,42	0,07	0,09	12,31	6,18

Table 4 Option Statistic Question No. 15

Opt	N	Prop.	Rpbis	Rbis	Mean	SD
A	8	0,07	-0,14	-0,27	8,63	2,00
B	18	0,17	0,08	0,11	12,28	5,32
C	10	0,09	-0,18	-0,31	8,40	4,40
D*	**70**	**0,65**	**0,19**	**0,24**	**13,11**	**5,66**
E	2	0,02	-0,21	-0,63	3,00	0,00

Table 5 Option Statistic Question No. 7

Opt	N	Prop.	Rpbis	Rbis	Mean	SD
A	3	0,03	-0,09	-0,22	8,67	2,89
B	24	0,22	-0,20	-0,27	9,42	3,28
C	10	0,09	-0,30	-0,52	6,40	2,41
D	0	0.00	--	--	--	--
E*	**71**	**0,66**	**0,38**	**0,50**	**13,83**	**5,73**

Figure 2 Data Group Graphic Distribution Vs Proportion Correct on Question No 1

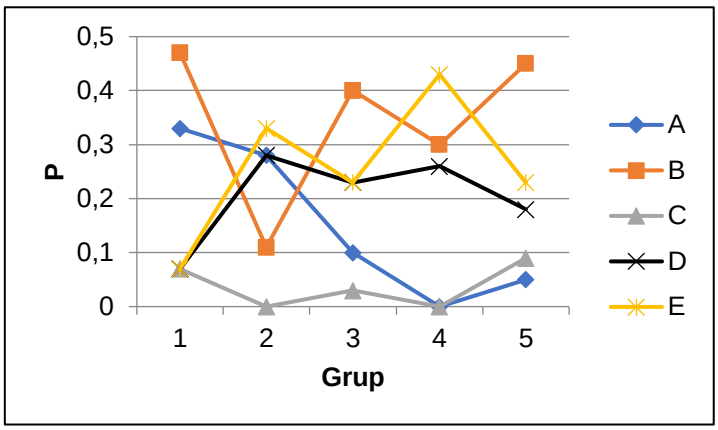


Figure 5 Data Group Distribution vs Proportion Correct Question No.18

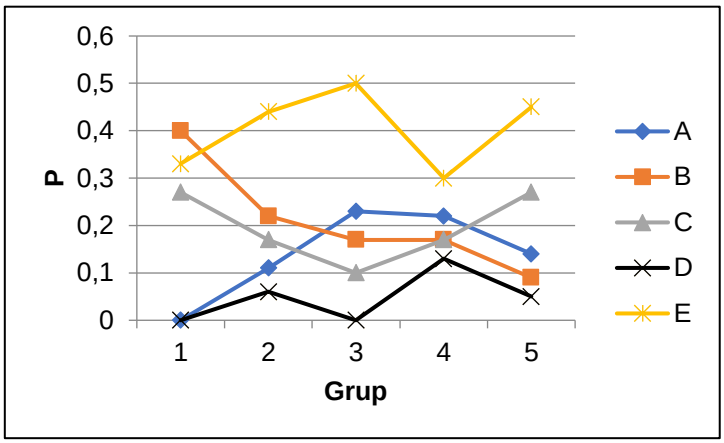
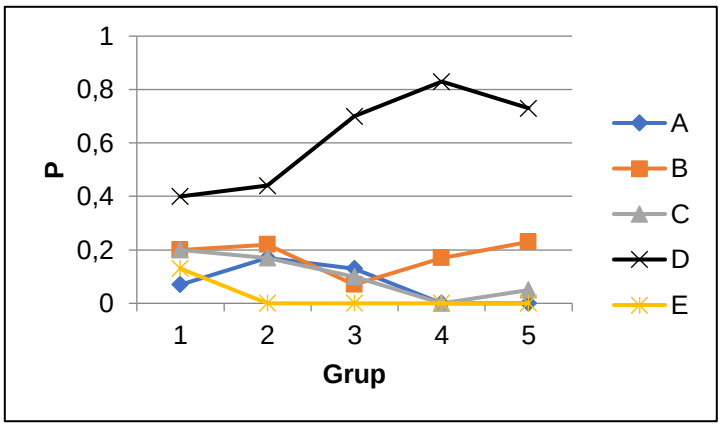


Figure 7 Data Group Distribution vs Proportion Correct Question No.15



Validity can also be interpreted as the accuracy of measuring instruments, such as construction validity, content validity, forecast validity, and equal current validity. If the r count is greater than the r table with a significance level or error level of 5% (r-hit > r-tab with a significance level of 5%), then the test item is valid.

The results of the validity analysis on the PAS Chemistry Class XI questions of High School 1 Wonogiri obtained 76.67% of valid questions and 33.34% of invalid questions.

One of the invalid questions is question number 01. In question number 1, the value of the *rpbis option* and *rbis option* in answer key Option E is 0. While answer option B has *rpbis option* and *rbis option* of 0.07 and 0.09, answer Option C has *rpbis option* and *rbis option* of 0.09 and 0.21, answer Option D has *rpbis option* and *rbis option* of 0.06 and 0.09. The *error code* K in question number 1 comes from answer Options B, C, and D because the value of *the rpbis option* and *rbis option* in the answer option is higher than the answer key in Option E.

Figure 1 Question No.1

Mengenai reaksi eksoterm,pernyataan berikut yang benar adalah….

A. entalpi awal lebih besar dari entalpi akhir dan $\Delta H > 0$
B. entalpi awal lebih kecil dari entalpi akhir dan $\Delta H < 0$
C. entalpi awal sama dengan entalpi akhir dan $\Delta H = 0$
D. entalpi awal lebih kecil dari entalpi akhir dan $\Delta H > 0$
E. entalpi awal lebih besar dari entalpi akhir dan $\Delta H < 0$

Table 1 shows that 29 students chose Option E (answer key), 38 students chose Option B, four chose Option C, and 23 chose Option D. This means that key E is chosen second most by students. Looking at Table 1, group 1 indicates the student with the lowest score, and group 5 indicates the student with the highest score. Group 2 students chose answer key E by 33% and group 5 by 23%. Then, students who chose Option B in Group 2 as much as 11% and Group 5 as much as 45%. Students who chose Option C in Group 2 as much as 0% and Group 5 as much as 9% and those who chose Option D in Group 2 as much as 28% and Group 5 as much as 18%. This means that many students with high scores are fooled into choosing answer Option B over answer Option E, so this causes an error with the K code on question 1. Question number 1 is a category C1 question in rote form about exothermic reactions. This reaction will release heat from the system to the environment because by releasing heat, the enthalpy value will decrease, and ΔH is negative. In answer options B and E, ΔH<0 or ΔH is negative, then look at the previous sentence. The enthalpy change or ΔH is the difference between the product entapi and the reactants so that in the exothermic reaction, the initial enthalpy is greater than the final enthalpy. Lack of understanding of theory makes students score high fooled at Option B. The lack of understanding of exothermic reaction material that refers to students' answers is one of the factors that cause invalidity.

A good answer key will show the slope of the line to the top right. In Figure 2, it can be seen that answer key E shows a downward slope line, which indicates that answer key E is not good because the downward slope indicates that only a few students scored five who chose answer key E. Student answers to the PAS question in this study affect the results of the analysis of Iteman 4.5. It can be concluded that the question's validity is influenced by student answers.

2. Reliability

According to Sudijono (2011), a test that is tried many times on the same subject always shows fixed results, and firm and stable tests can be said to be reliable [18]. The word reliable itself means trustworthy. The Alpha Cronbach, Kuder-Richardson (KR-20 or KR-21) Coefficient formulas and the Halving Technique can be used to measure the reliability of internal consistency. The problem is said to have a fairly high reliability if it has a reliability coefficient between 0.7-0.8 based on the Kuder-Richardson formula. At the same time, Alpha Cronbach can be used to analyze question items by giving false scores of 0 and true +1 or with scores of 1, 2, and

3 consecutively. This method is a step to determine the reliability coefficient of the instrument/test, which refers to the concept of internal consistency.

The internal consistency method is based on the correlation between answer scores on each question item or homogeneity. If the correlation is close to zero, then the internal consistency is zero, and the reliability is low. The reliability is also high if the average correlation between question items is high [1].

Table 2. Reliability Criteria

Benchmark	Criteria
<0,20	Very Low
0,20 - 0,40	Low
0,40 – 0,60	Medium
0,60 – 0,80	High
0,80 – 1,00	Very High

The reliability analysis results on the PAS Chemistry Class XI High School 1 Wonogiri obtained an alpha reliability value of 0.82 and was in the high category.

3. Difficulty Level

The difficulty or difficulty of a question item shows how likely it is for respondents to answer an item correctly. Difficulty category [20]:

1) P value < 0.3 can be interpreted as a question of entering the difficult category
2) P value between 0.3-0.7 can be interpreted as a question of entering the medium category
3) P value > 0.7 can be interpreted as a question in the easy category

If the difficulty level is more significant, then the problem is more manageable, and vice versa. The difficulty level of a question item does not indicate that a particular question item is good or bad but only indicates that it is difficult or easy for a specific group of test takers. Question items that are too difficult or too easy do not provide much information about the question item or test taker. Several ways to express difficulty according to classical theory include linear difficulty scales, Davis indexes, bivariate scales, and proportions of correct answers.

The results of the difficulty level analysis on the PAS Chemistry Class XI questions of High School 1 Wonogiri obtained 33.34% of the difficult category questions and 66.66% of the medium category questions.

Figure 3 Question No. 8

Kenaikkan suhu umumnya menaikkan laju reaksi. Alasan yang tepat untuk menjelaskan hal itu adalah….

A. Energi kinetik dari molekul-molekul menurun
B. Kenaikkan suhu menghasilkan reaksi dapat balik
C. Kecepatan masing-masing molekul sama
D. Energi kinetik dari molekul-molekul meningkat
E. Kenaikkan suhu memperkecil energi aktivasi

One of the questions that falls into the medium category is question number 8. The result of Iteman 4.5, question number 8, has a proportion correct of 0.7 or can be interpreted as many as 70% of students choose the answer key option, and the other 30% choose another answer option. This number makes the results of Iteman 4.5 analysis on question 8 categorized as medium. Question 8 is a question about one factor that affects the reaction rate, namely, temperature increase. According to Nazar (2010)An increase in temperature in a chemical reaction can increase the kinetic energy of the reacting substances so that the reaction takes place faster so that the correct and appropriate answer is the answer. Option D is the answer key.

Furthermore, the question that falls into the difficult category is question number 11. This problem is also a C2 problem because it tests the ability to understand and interpret the reaction rate material. In question number 11, it is known that the correct proportion is 0.29, or it can be interpreted that only 31 students chose the answer key contained in Option E. This number makes question number 11 categorized as a difficult question.

Figure 4 Question No. 11

Diketahui data hasil percobaan untuk reaksi P + Q sebagai berikut :

No	Zat yang bereaksi		Waktu (detik)	Suhu (^oC)
	P	Q		
1	1gram serbuk	2 M	20	26
2	1gram larutan	2 M	10	26
3	1gram kepingan	2 M	40	20
4	1gram serbuk	4 M	5	26
5	1gram larutan	4 M	5	30

Faktor yang memengaruhi laju reaksi pada percobaan 1 dan 3 adalah….

a. suhu
b. konsentrasi
c. luas permukaan dan konsentrasi
d. Suhu dan luas permukaan
e. luas permukaan dan suhu

Question 11 is question C4, which is a question that applies and analyzes relationships or factors that affect reaction rate. This question asked what factors influenced experiments no. 1 and 3. Substance P in experiment 1 has a mass of 1 gram in powder form, while substance P in experiment 3 has a mass of 1 gram in pieces. Then the concentration of substance Q in experiments 1 and 3 is the same, which is 2M. Then, the time of Experiment 1 was 20 seconds shorter than Experiment 3, and the temperature of Experiment 1 was 6 degrees Celsius higher than Experiment 3. The reaction rate factors that affect experiments 1 and 3 are surface area and temperature. This means that in the question above, answer Option E is correct because experiments 1 and 3 begin with differences in surface area and continue with temperature differences. The proportion of students who chose other options as answers, more than 50% of the total students, showed that many students needed help understanding what factors affect the reaction rate if given a comparison model in the experiment.

4. Discriminatory Power

Discriminatory power analysis is an activity carried out to examine question items by distinguishing students' abilities based on their high and low achievements. The higher the distinguishing power value of a question item, the better the question item is at distinguishing the ability of high and low-score students. Discriminatory power criteria and discriminatory power interpretation [18]:

1) The power of difference value of < 0.0 is included in the criteria for a terrible question. Can be interpreted as discarded question items
2) The power of difference value of 0.0-0.20 is included in the criteria for bad questions. Can be interpreted as discarded question items
3) The power of difference value of 0.21-0.40 is included in the criteria for medium questions. Can be interpreted as revised question items
4) The power of difference value of 0.41-0.70 is included in the suitable criteria. Can be interpreted with little or no revision of the question items
5) The power of difference value of 0.71-1.00 is included in the criteria

for an excellent question. It can be interpreted that question items work well

The results of the power of difference analysis on the PAS Chemistry Class XI questions of High School 1 Wonogiri obtained 7% of the questions in the terrible category, 20% of the questions in the bad category, 40% of the questions in the medium category, and 33% of the questions in the excellent category.

Question 18 is a question with a very bad category. This problem has an rpbis value of -0.01. The answer key is located in Option A. In Figure 4. 15, you can see the gradient of Option A heading down. This indicates that answer key A is not widely chosen by group 5 students.

Question 18 is a question with a very bad category. This problem has an rpbis value of -0.01. Question number 18 has a negative RPBIS value. Compared to other answer options where the rpbis value is positive in answer options C, D, and E and the rpbis value is negative in answer Option B, there is an LR error code. An inferior power of difference value indicates that the question item cannot distinguish high-ability students from low-ability students; a total negative value of RPBIS indicates it. In other words, a small RPBIS score indicates that a high score or competent student chooses another answer option.

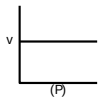
Question 18 is question C4, which tests the understanding of the graph depicting the reaction rate to R. To find out, first find order X by comparing experiments 1 and 3. After that, order X is 0, so the graph that illustrates the level of reaction to R is a graph with horizontal lines in answer Option A.

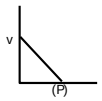
Figure 6 Question No. 18

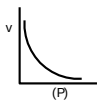
Data untuk reaksi : R + Q $\rightarrow$ hasil adalah sebagai berikut :

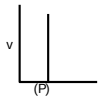
No	[R] (mol/dt)	[Q] (mol/dt)	Laju (mol/Ldt)
1	0,1	0,1	$2,2 \times 10^{-4}$
2	0,2	0,3	$19,8 \times 10^{-4}$
3	0,1	0,3	$19,8 \times 10^{-4}$

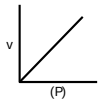
Maka grafik yang menggambarkan tingkat reaksi terhadap R adalah….

A. 

B. 

C. 

D. 

E. 

Question 15 includes questions with bad categories. This question has an rpbis value of 0.19. The answer key is located in Option D. Question number 15 has a more excellent rpbis value in the key than other negative answer options. The rpbis value of answer options A, C, and E has a minus value where students with low ability choose the answer option, and the rpbis value of Option B has a positive value where students with high ability choose the answer option.

Suppose you look at Table 4 and Figure 7. In that case, the rpbis value of answer Option D is positive, and the gradient of Option D is pointing down, which indicates that few high-ability students answered correctly question number 15. In contrast, options A, C, and E show negative RPBIS values and a downward slope, which indicates that many low-ability students choose the

option, while the positive RPBIS Option B and the upward gradient of Option B indicates that many high-ability students are deceived into choosing the answer to Option B.

5. The Effectiveness of The Deceiver

The function of the deception effectiveness analysis is to determine the extent to which the student chooses other answer options besides the key. Deceivers or distractors in multiple-choice questions are categorized as good or successful if students with low scores prefer deceivers compared to high-score students [14]. Fraudsters should be selected with a minimum limit of 5% of learners or students, and the creation of deceivers should be similar to answer keys.

Each distractor has a definite value of good and not good. Therefore, the Likert scale was used in this study to determine the effectiveness of distractors in the problem:

1) The number of deceivers is 0, it can be interpreted as a terrible problem deceiver
2) The number of deceivers is 1, which can be interpreted as a deceiver about bad questions
3) The number of deceivers is 2, which can be interpreted as enough problem deceivers
4) The number of deceivers is 3, which can be interpreted as a good problem deceiver
5) The number of deceivers is 4, which can be interpreted as very good problem deceivers

The results of the analysis of the effectiveness of deceivers on the PAS Chemistry Class XI questions of High School 1 Wonogiri obtained 7% of the questions in the sufficient category, 30% of the questions in the excellent category, and 63% of the questions in the outstanding category.

Figure 8 Question No. 7

Semakin tinggi konsentrasi zat-zat pereaksi, reaksi berlangsung semakin cepat. Hal ini dikarenakan semakin….

A. Rendah energi pengaktifan reaksi
B. Besar energi yang dihasilkan partikel
C. Cepat gerakan antar partikel
D. Pendek jarak antar partikel
E. Banyak kemungkinan partikel bertumbukan

Question 7 is question C2, which is about understanding one of the reaction rate factors. This question is also a question with sufficient deceptive effectiveness criteria, where only two answer options managed to deceive by being chosen by more than 5% of students. Table 5 shows that the endorsing prop value in answer Options B and C is more than 0.05, so both options successfully deceive students. The rpbis value (in Table 5) of question number 7 of 0.38 shows that question number 7 has valid question validity, difficulty level, medium, and medium power of difference value. Question number 7 is a question about the reaction rate, where if the reagent concentration is higher, the reaction is faster. This is because the more particles that react, the more frequent collisions between particles occur. According to [18], if the deceiver has not functioned correctly, then the deceiver on the question should be corrected or replaced with another deceiver, which means that answer options A and D in question number 7 must be corrected or replaced with other answer options.

CONCLUSION

Based on the analysis of the quality of the Final Semester Assessment (PAS) questions for the Class XI Chemistry subject of High School 1 Wonogiri for the 2022/2023 academic year using Iteman 4.5, it can be concluded that:

1. In the validity aspect, 76.67% of the questions were valid, and 33.34% were invalid.
2. In the reliability aspect, an alpha reliability value of 0.82 was obtained and was in the high category.
3. Regarding difficulty, 33.34% of the questions were obtained in the difficult category, and 66.66% were in the medium category.
4. In the aspect of discriminatory power, 7% of the questions were very bad, 20% of the questions were in the bad category, 40% of the questions were in the medium category, and 33% of the questions were in the excellent category.
5. In the aspect of deceptive effectiveness, 7% of the questions were categorized as sufficient, 30% of the questions were classified as good, and 63% of the questions were categorized as very good

ACKNOWLEDGMENTS

Thank you to the principal and teachers of High School 1 Wonogiri for allowing conducting research.

REFERENCE

[1] Perwira Negara, H. R., & Kurniawati, K. R. A. (2018). Karakteristik Butir Soal Try Out Ujian Nasional SMA Mata Pelajaran Matematika Se-Kota Mataram dengan Menggunakan Program Iteman. Jurnal VARIAN, 1(2), 11–21. https://doi.org/10.30812/varian.v1i2.67

[2] Pristiwanti, D., Badariah, B., Hidayat, S., & Dewi, R. S. (2022). Pengertian Pendidikan. Jurnal Pendidikan dan Konseling (JPDK), 4(6), 1707–1715. https://doi.org/https://doi.org/10.31004/jpdk.v4i6.9498

[3] Arifin, Z. (2015). Evaluasi Pembelajaran Penulis. Direktorat Jenderal Pendidikan Islam Kementerian Agama RI.

[4] Erfan, M., Maulyda, 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://ejournal.undiksha.ac.id/index.php/IJERR/article/view/24080/pdf

[5] Santosa, S., & Badawi, J. A. (2022). Analisis Butir Soal Pilihan Ganda Tema Pertumbuhan dan Perkembangan Makhluk Hidup Kelas III Madrasah Ibtidaiyah. Jurnal Basicedu, 6(2), 1678–1686. https://doi.org/10.31004/basicedu.v6i2.2206

[6] Prabowo, A., Sunaryo, S., & Rahmawati, U. (2017). Pengembangan Modul Analisis Butir Soal dengan Menggunakan Item and Test Analysis. AdMathEdu : Jurnal Ilmiah Pendidikan Matematika, Ilmu Matematika dan Matematika Terapan, 7(2), 99. https://doi.org/10.12928/admathedu.v7i2.9146

[7] Nayla Amalia, A., & Widayati, A. (2012). Analisis Butir Soal Tes Kendali Mutu Kelas XII SMA Mata Pelajaran Ekonomi Akuntansi di Kota Yogyakarta. Jurnal Pendidikan Akuntansi Indonesia, X(1), 1–26. https://doi.org/http://dx.doi.org/10.21831/jpai.v10i1.919

[8] Hadiana, D. (2015). Penilaian Hasil Belajar untuk Siswa Sekolah Dasar. Jurnal Pendidikan dan Kebudayaan, 21(1), 15. https://doi.org/10.24832/jpnk.v21i1.173

[9] Subali, B. (2014). Analisis Soal Baik Kualitatif Maupun Kuantitatif. Workshop Item Development Dosen

Poltekes Kebidanan Politeknik Kesehatan Surakarta, 5, 1–36.

[10] Fatimah, L. U., & Alfath, K. (2019). Analisis Kesukaran Soal, Daya Pembeda dan Fungsi Distraktor. Al-Manar, 8(2), 37–64. https://doi.org/10.36668/jal.v8i2.115

[11] Nazar, M. S. S. W. R. F. et al. (2010). Identifikasi Miskonsepsi Siswa SMA pada Konsep Faktor-Faktor yang Mempengaruhi Laju Reaksi. Jurnal Biologi Edukasi, 2(3), 49–53.

[12] Jihad, Asep dan Haris, Abdul. (2013). Evaluasi Pembelajaran. Yogyakarta: Multi Pressindo

[13] Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan Rasch pada Assessment Pendidikan. In Cimahi Penerbit Trim Komunikata.

[14] Daryanto. (2012). Evaluasi Pendidikan. Jakarta: PT Rineka Cipta

[15] Perdana, S. A. (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

[16] Drost, E. A. (2011). Validity and Reliability in Social Science Research. Education Research and Perspectives, 38(1), 105–123.

[17] Cohen−Swerdlik. (2009). Psychological Testing and Assessment: An Introduction to Tests and Measurement 7th Edition. McGraw-Hill.

[18] Sudijono, A. (2015). Pengantar Evaluasi Pendidikan, cet. Ke-14. Jakarta: Raja Grafindo Persada

[19] Guilford, J.P. (1956). Fundamental Statistic in Psychology and Education 3rd Ed. New York: McGraw-Hill Book Company

[20] Triono, D., Sarno, R., & Sungkono, K. R. (2020). Item Analysis for Examination Test in The Postgraduate Student's Selection with Classical Test Theory and Rasch Measurement Model. Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020, 523–529. https://doi.org/10.1109/iSemantic50169.2020.9234204