MALAYSIAN JOURNAL OF INNOVATION IN ENGINEERING AND APPLIED SOCIAL SCIENCE (MyJIEAS)
e-ISSN : 2811-3780
VOLUME 41, ISSUE 01

Analysis of Final Examination Paper of Chemistry 1 Course for Polytechnic TVET Foundation Program Using Rasch Measurement Model

S.N.A. Aziz[1], N. Hairi[1]

[1]Examination and Evaluation Division, Department of Polytechnic and Community College Education, 6200 Putrajaya, Malaysia.
Corresponding Author's Email: [1]nadia.hairi@mohe.gov.my

Article History : Received 10092024; Revised 28092024; Accepted 15102024;

ABSTRACT – This study aims to analyze the final examination paper of Chemistry 1 Course using Rasch Measurement Model. This course is a mandatory component for all students enrolled in the TVET Foundation program at the Polytechnic. For the purposes of this study, two polytechnics were randomly selected, resulting in a sample size of 175 students. The examination paper comprises five structured questions amounting to a total of 22 items. The study employed the Rasch model through Winsteps software to conduct a descriptive analysis, assessing the reliability of item constructs. This included calculating the reliability index to ensure measurement consistency, the separation index, item polarity, unidimensionality, item fit, and item difficulty level. The results indicated that the reliability index was 0.97, signifying a high level of reliability and consistency in the developed items. However, some items were identified as needing refinement, particularly those classified as moderate and difficult, which posed challenges for students' performance. The final analysis found out that the set of questions presented adhered the psychometric standards established by the Rasch model.

KEYWORDS: Reliability; TVET; Rasch Measurement Model

1.0 INTRODUCTION

Polytechnics and Community Colleges are key TVET (Technical and Vocational Education and Training) institutions contributing significantly to the skilled and semi-skilled workforce needed in the national industry. According to the Deputy Prime Minister, who is also the Chairman of the National TVET Council, during the National TVET Day on June 2, 2023, the Fourth Industrial Revolution and the digital era require the younger generation to equip themselves with new skills to compete on the global stage. In response to this, JPPKK (Department of Polytechnic and Community College Education) is advancing the national agenda to empower TVET education through extensive collaboration with the industry and partnerships between the government and GLCs (Government-Linked Companies) to produce a skilled workforce that will contribute to national development. The necessity of skills through TVET education is widely acknowledged as it forms a part of the national agenda to meet current job market demands.

The Polytechnic Transformation 2023 – 2030 is a continuous effort to enhance the quality of Polytechnic graduates in line with the current national needs. Among the emphasized agendas in this transformation plan is ensuring that the study programs offered remain relevant to industry requirements and meet the standards set by accreditation bodies as a guarantee of educational quality. The newly introduced TVET Foundation Program in 2023 aims to widen access for Malaysian Certificate of Education (SPM) leavers to pursue preparatory programs before advancing to undergraduate studies. The Engineering Technology Foundation Program, previously known as the TVET Foundation Program, is the pioneering program offered at five selected polytechnics for outstanding SPM graduates. This program serves as a special pathway preparing students through a foundation program before proceeding to undergraduate studies at the Malaysian Technical University Network (MTUN), which includes UTeM, UNIMAP, UMPSA, and UTHM. JPPKK has established collaboration with MTUN, representing four leading TVET universities that will welcome TVET Foundation graduates to continue their studies in the Bachelor of Engineering Technology programs based on the entry requirements set by each university.

The Rasch Model is a popular model used among researchers aiming to analyze the reliability of developed items. This model facilitates identifying item validity levels, and item construct reliability, ensuring the developed items match the stated difficulty levels, and guaranteeing the achievement of learning outcomes stated in a course curriculum.

2.0 RESEARCH OBJECTIVES

Among the criteria qualifying foundation students to proceed to MTUN is achieving a minimum Cumulative Grade Point Average (CGPA) of 2.00 and passing all courses taken during the TVET Foundation Program. The first cohort of students participated in this program in Session 1:2023/2024, with a capacity of 500 students nationwide. Based on the Semester 1 examination results records obtained by the Examination and Evaluation Division, JPPKK, for all foundation students, it was found that certain courses contributed to low CGPA achievements. Specifically, the course FB10054: Chemistry 1 had only 15% of students passing with a minimum score of 40 marks. The institution has provided feedback indicating that the low scores resulted from the developed questions being too difficult and not matching the foundation students' level. Therefore, BPN conducted a study to analyze the scores obtained by students for each item in the final exam for Chemistry 1 using the Rasch Model to identify the reliability of the developed items. The analysis is based on several criteria such as item-individual reliability and separation index; item-individual separation index; item polarity; unidimensionality; item fit; and item difficulty level.

3.0 LITERATURE REVIEW

This paper aims to identify the validity level of items developed for the final exam for students taking Chemistry 1 in the TVET Foundation Program at Polytechnics. The analysis is carried out by taking the marks for each item answered by students and testing the reliability, separation index, and item fit analysis through the distribution of marks obtained by students based on each item. The Rasch Model is one suitable tool for analyzing these items. According to [1], the Rasch model measurement framework is well-established in the education field due to its ability to validate instrument construction, quality analysis, and individual achievement analysis through measurement yielding empirical results. The probability of an individual correctly answering a test question can be identified, as well as the difficulty level of the given question.

Most tests using the Rasch Model focus on multiple-choice questions rather than subjective questions. However, this study on the final exam paper for Chemistry 1 focuses on the marks obtained by students for subjective questions. The analysis can still be performed to test the credibility of the questions and students through the obtained scores. According to [2], the Rasch measurement model provides good psychometric item analysis to determine the quality of each item. Items that cannot differentiate between students are unsuitable for measuring student abilities and should be discarded or modified.

The Rasch Model has several key principles developed by previous researchers in the 1960s, where the probability of a student correctly answering a given question depends on the difficulty level of the question and the student's ability. This indicates that an important feature of this model is its ability to provide a range reading for both aspects: the used items and the knowledge level of the students answering the questions. The model assumes that highly capable individuals are more likely to answer items correctly, while easy items can be answered correctly by all respondents when controlled by the difference between item difficulty and respondent ability [3] .

The item development process for the final exam is overseen by the Examination and Evaluation Division, JPPKK, following the rules and procedures outlined in the Polytechnic Item Bank Management Guidelines and Marking Regulations. Item development involves subject lecturers and is reviewed and validated by appointed expert lecturers. The marks obtained for each item can also be analyzed to identify the validity and reliability of the items used for the final exam. The advantage of using the Rasch Model is its ability to identify the suitability of items and

students answering them, allowing this method to determine the question difficulty level and student ability level [4]. The Rasch Model provides a robust and elegant framework for measuring individual abilities or attributes considering the characteristics of the items used in the measurement.

4.0 METHODOLOGY

The final examination paper for the TVET Foundation Program offered at polytechnics is a standardized and centralized set of questions. The control and management of each item's development fall under the jurisdiction of the Examination and Evaluation Division, JPPKK. Lecturers at the institutions are appointed as question setters based on their subject matter expertise. The examination schedule for each course, including the dates and times for the final exams, has been standardized to prevent any leakage of exam questions, as all institutions receive the same set.

The document utilized for this study is the final examination paper for the course of Chemistry 1. The examination paper comprises five structured questions, each containing multiple sub-items, amounting to a total of 22 items. The questions included in this examination are standardized covering every topic specified in the Final Examination Item Specification Table (FEIST). The FEIST development refers to the difficulty levels and learning outcomes set within the curriculum.

This final examination paper administered across five polytechnics in Malaysia. However, for the purposes of this study, a random selection process was employed to choose two polytechnics, which is Politeknik Ungku Omar (PUO) and Politeknik Merlimau Melaka (PMM), resulting in a sample size of 175 students.

For data preparation, each student's score for every question was recorded in Microsoft Excel, including detailed records for each question with sub-items. The complete score records were then transferred into the Winsteps software for data analysis. The examination paper was analysed using Winsteps software to identify the reliability of the constructed items. The analysis performed was (i) reliability index; (ii) separation index; (iii) items polarity; (iv) unidimensionality; (v) items fit and (vi) level of difficulty of items. Table 1 below shows the setting as a guide for each analysis using the Rasch Measurement Model.

Table 1: Rasch Model Setting

Analysis	Setting	Source
Reliability index	> 0.8	Bond & Fox (2015)
Separation index	> 2.0	Bond & Fox (2015)
Items polarity (PMC)	> 0.0	Linacre (2012)
Unidimensionality		
Raw variance explained by measures	> 40%	Abdul Aziz et al. (2017)
Unexplained variance in 1st contrast	< 15%	Abdul Aziz et al. (2017)
Eigen	< 5	Linacre (2005)
Items fit		
Infit MNSQ	0.5 – 1.5	Bond & Fox 2015)
Outfit MNSQ	0.5 – 1.5	Bond & Fox 2015)

5.0 RESULTS AND FINDINGS

5.1 Analysis of Reliability and Separation Items and Respondent

Analysis of reliability index is important to ensure the consistency in measuring what should be measured. The level of reliability of the study can be determined by interpretating the Cronbach's Alpha value which ranges between 0 to 1.0. A range value near to 1.0 indicates that the level of reliability is at a level with high consistency. A range value near to 0 indicates a low

level of reliability. Table 2 below shows the results of the analysis of reliability and separation index for items and individuals.

Table 2: Analysis Of Reliability and Separation Index

	Reliability index	Separation index
Item	0.97	5.67
Individual	0.76	1.77

Based on the data presented in Table 2, the reliability index of the item is 0.97, indicating a very high level of reliability and consistency. The item separation value is 5.67, demonstrating that the item is well-calibrated, allowing the Rasch Model to consistently identify five distinct levels of item difficulty. According to [6], a separation index greater than 2.0 signifies a good distribution between items and individuals.

The individual reliability index is 0.76, which reflects a good and acceptable level of reliability [5]. The individual separation index is 1.77, which, when rounded to 2.0, remains acceptable. This indicates that the Rasch Model can consistently detect nearly two levels of individual ability clusters.

5.2 Analysis of Item Polarity

Item polarity analysis is conducted to ensure that all items align in one direction to effectively measure the intended construct. This analysis involves examining the Point Measure Correlation (PTMEA CORR.) values. According to [5], a positive PTMEA CORR. value indicates that the item measures the intended construct, whereas a negative value suggests that the item does not measure the intended construct. In this study, the PTMEA CORR. values were all positive, ranging from 0.23 to 0.57, indicating that all items successfully measure the intended construct.

Table 3: Point Measure Correlation Value

Entry Number	Item	PT-MEASURE CORR.	Entry Number	Item	PT-MEASURE CORR.
22	5cii	0.32	3	1b	0.48
20	5b	0.27	15	4a	0.51
21	5ci	0.28	1	1ai	0.23
19	5a	0.45	8	2ci	0.49
17	4ci	0.27	16	4b	0.57
18	4cii	0.45	6	2a	0.56
13	3ci	0.49	5	1cii	0.54
12	3bii	0.30	9	2cii	0.45
11	3bi	0.31	7	2b	0.50
2	1aii	0.50	4	1ci	0.48
14	3cii	0.42	10	3a	0.55

5.3 Unidimensional

Unidimensionality is a crucial characteristic for ensuring that items in a measuring instrument align in one direction. Unidimensional analysis of the items was conducted following the assessment of their polarity values. To confirm that items exhibit strong unidimensional properties, the raw variance explained by the measures should be at least 40%, and the unexplained variance in the first contrast should not exceed 15% [7]. Additionally, the eigen value should be less than 5 [6], and the minimum ratio of unexplained variance in the first contrast to the raw variance explained by measures should be at least 3:1 [8], [9].

Table 4: Analysis of Unidimensional

	Empirical	Model
Raw variance explained by measures	42.6%	43.4%
Unexplained variance in 1st contrast	6.7%	
Eigen value	2.6	
Ratio	6:1	

The findings indicate that the raw variance explained by the measures is 42.6%, compared to the Rasch model which is 43.4%. The value of unexplained variance in 1st contrast is at a rate of 6.7%. The eigen value of 2.6 shows that there is no second dimension [6] while the value of the ratio of unexplained variance in 1st contrast with the value of raw variance explained by measures is at a rate of 6:1. These results collectively demonstrate that the developed item is unidimensional.

5.4 Item Fit Measure

Item fit analysis is conducted to determine whether the constructed items exhibit the characteristics of consistency as defined by the Rasch Measurement Model. The acceptable range for item fit is based on the mean square (MNSQ) values for both infit and outfit, which should fall between 0.5 and 1.5 [5]. A high infit MNSQ value above 1.5 suggests that student responses are inconsistent, indicating a mismatch with the Rasch model. Similarly, a high outfit MNSQ value above 1.5 indicates that students are confused by the question, while a value below 0.5 suggests that the question is too easy. An item is considered misfitting if its z-standardized (z-std) index exceeds the threshold of +/-2 logits; however, this index can be disregarded if the MNSQ infit and outfit values are within the acceptable range [10]. Items failing to meet these criteria should be reviewed for potential refinement or removal. Table 5 below lists the items that fall outside the acceptable MNSQ range of 0.5 to 1.5.

Table 5: Misfit Item

Item	Infit		Outfit		Decision
	MNSQ	z-std	MNSQ	z-std	
5a	1.81	2.4	0.60	-1.4	Dropped
4a	1.54	4.7	1.49	3.7	To be reviewed

Based on the table above, there are two (2) items which is item 4a and item 5a that do not fall within the acceptable range for item fit. Item 4a, with an infit MNSQ value of 1.54, is slightly above the threshold, but its outfit MNSQ value of 1.49 is acceptable. Therefore, it is recommended to retain item 4a with some refinement. In contrast, item 5a has a high infit MNSQ value of 1.81, indicating a significant inconsistency between student responses and the Rasch model. Consequently, it is recommended that item 5a be removed from the assessment.

5.5 Analysis of item difficulty level

The difficulty level of an item can be assessed using individual-item maps. These maps illustrate the relationship between students and items, providing a visual representation of the distribution of item difficulty, ranging from easy to difficult.

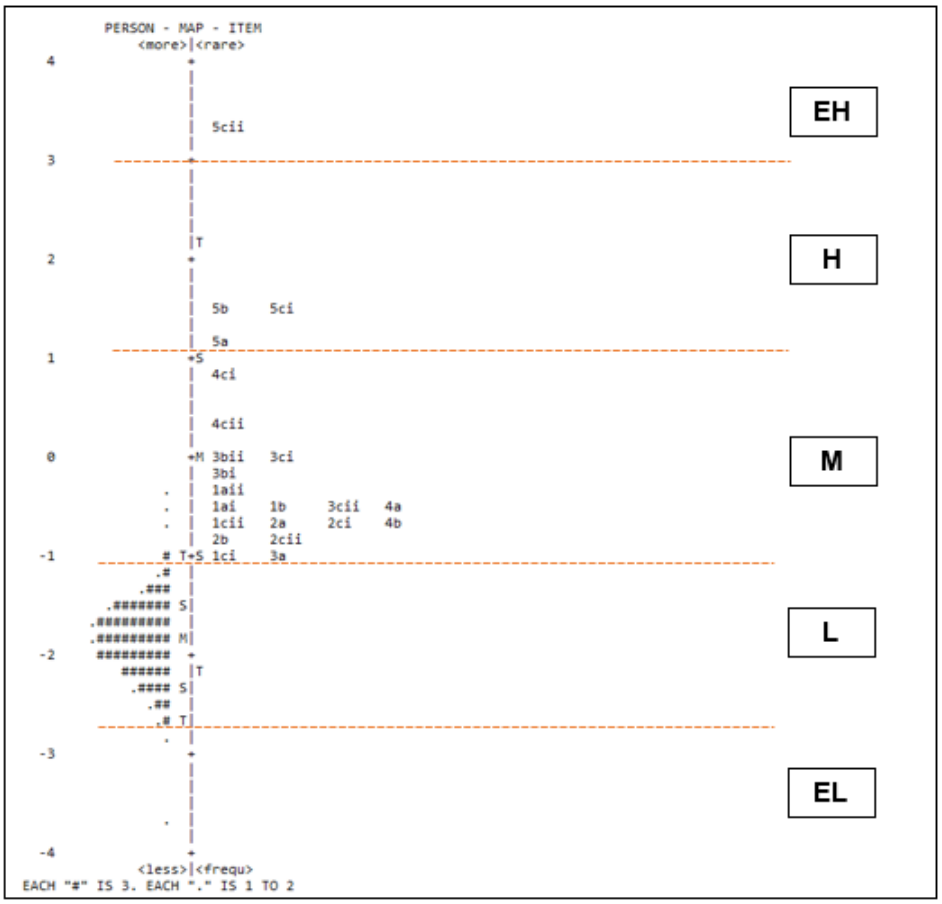

Figure 1. Person-item map

Based on Figure 1, item 5cii, with a logit value of +3.47, is the most difficult for students to answer. There are three (3) items in the difficult level which are items 5a, 5b and 5ci while the rest of the items are easier items that can be answered by students. Items 1ci and 3a, both with a logit value of -1.0, are the easiest for students, despite being classified as medium difficulty. Most students are positioned at the -1 logit level, indicating a generally low ability to answer the questions, likely due to the high difficulty of the test items. This is evident from the relative positioning of items and student abilities on the person-item map.

5.6 Summary of Analysis

The analysis of the final examination paper for the Chemistry 1 subject was conducted using the Rasch Measurement Model, involving 22 items and 175 students. The analysis suggested the removal of one item during the item matching process, while another item requires refinement based on feedback from the item drafter. The remaining items were retained in accordance with the assumptions of the Rasch Measurement Model. However, the concentration of items at medium and difficult levels may have contributed to students' challenges in performing well. The most difficult items should be reviewed and adjusted to better align with students' abilities. Table 6 below provides a summary of the analysis conducted.

Table 6: Summary of Analysis Using Rasch Measurement Model

Analysis	Findings	Remarks
Item reliability	0.97	-
Individual reliability	0.76	-
Item separation index	5.67	-
Individual separation index	1.77	-
Item polarity (PMC)	0.23 – 0.57	-

Unidimensionality		
Raw variance explained by measures	42.6%	-
Unexplained variance in 1st contrast	6.7%	-
Eigen	6:1	-
Item fit		
Infit MNSQ	0.43– 1.46	Item 5a dropped
Outfit MNSQ	0.59– 1.49	-

6.0 CONCLUSION

In conclusion, the findings indicate that the items in the final exam paper for Chemistry 1 meet the psychometric standards set by the Rasch Measurement Model. Furthermore, the findings of this study provide lecturers with valuable information on the quality of exam items and the abilities of students based on their response patterns. Additionally, these insights help lecturers identify the strengths and weaknesses of students in specific topics represented by each item in the exam paper. This enables lecturers to plan appropriate approaches to improve students' understanding of the topics studied, which in turn, will directly contribute to better CGPA achievements for the students.

REFERENCES

[1] M. A. K. Mohd Lot and A. Z. Khairani, “Analisis Model Rasch Penilaian Akhir Bahasa Arab Kolej Vokasional,” *Malaysian Journal Of Social Sciences And Humanities (MJSSH)* , vol. 7, no. 10, 2022.

[2] M. E. E. Mohd Matore and S. Sovey, “Penilaian Tahap Kesukaran Item Ujian Akhir Sesi Akademik (UASA) Matematik Tingkatan 1,” *International Conference On Business Studies And Education (ICBE)*, 2023.

[3] A. Basran and D. Lajium, “Aplikasi Model Rasch Dalam Pengujian Instrumen Inventori Konsep Daya,” *International Journal Of Modern Education (IJMOE)*, vol. 2, no. 6, pp. 14–27, 2020.

[4] R. Mohd Yasin, M. F. A. N. Yunus, R. Che Rus, A. Ahmad, and M. B. Rahim, “Validity And Reliability Learning Transfer Item Using Rasch Measurement Model ,” in *4th World Congress On Technical And Vocational Education And Training (WOCTVET)*, Malaysia, 2014.

[5] T. G. Bond and C. M. Fox, *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*, 3rd ed. New York: Routledge, 2015.

[6] J. M. Linacre, *Test Validity and Rasch Measurement: Construct, Content, etc.* 2005.

[7] A. Abdul Aziz, M. S. Masodi, and A. Zaharim, *Asas Model Pengukuran Rasch : Pembentukan Skala dan Struktur Pengukuran*, Edisi 3. Bangi: Penerbit UKM, 2017.

[8] K. J. Conrad *et al.*, “Dimensionality, Hierarchical Structure, Age Generalizability And Criterion Validity of The GAIN’S Behavioral Complexity Scale,” *Psychol Assess*, vol. 24, pp. 913–924, 2012.

[9] S. Embretson and S. Reise, *Item Response Theory For Psychologists*, vol. 4. Mahwah, NJ: Erlbaum, 2000.

[10] M. Md Said, Z. Haron, and S. Surat, “Kesahan dan Kebolehpercayaan Instrumen Literasi Pentaksiran Bilik Darjah Melalui Model Pengukutan Rasch,” in *Seminar Antarabangsa Isu-isu Pendidikan (ISPEN2018)*, Bangi. Selangor., 2018, pp. 24–35.