

# Test Instrument Validation with Rasch Model

Ayi Darmana<sup>1</sup>, Ahmad Nasir Pulungan<sup>2</sup>, Haqqi Annazili Nasution<sup>3</sup>, Mutiara Agustina Nst.<sup>4</sup>, Putri Faradilla<sup>5</sup>

[ayidarmana2013@gmail.com](mailto:ayidarmana2013@gmail.com)<sup>1</sup>

The Educational Chemistry Study Program, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan, Medan, North Sumatera, 20221, Indonesia<sup>1,3,4</sup>.

The Chemistry Study Program, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan, Medan, North Sumatera, 20221, Indonesia<sup>2,5</sup>

**Abstract.** One successful learning indicator is the value of student learning outcomes. To get this value, a learning achievement test instrument is given. The test instrument previously needed to be validated to produce accuracy between the score data on learning outcomes and learning objectives. Validation is also needed to improve the quality of the test instrument. Analysis techniques based on the Rasch model can be used to evaluate test instruments before being given to students. This research aims to determine the quality of the questions by validating the test instruments with the Rasch Model. The sample in this research was 29 students. The Students used as samples were from the Chemistry Department, Universitas Negeri Medan. The analysis results show that the test instrument quality is categorized as high at 0.91 in terms of item reliability. The validated test instrument consisted of 25 multiple-choice questions and 18 questions were declared valid.

**Keywords:** Rasch Model, Reliability, Validity, Winstep

## 1 Introduction

One of the indicators of successful learning is the acquisition of satisfactory scores or scores from learning outcomes. Learning outcomes are determined by assessing a learner's ability indirectly by looking at how they react to various questions or stimuli [1]. In addition, learning evaluation is required to measure the success of learning activities [2]. The process of gathering the data and information required to assess learning is known as learning assessment and how learning has proceeded so that the necessary assessments and improvements can be made to maximize the results. In addition to being required to have the ability to evaluate the learning process, educators are also required to have the ability to develop evaluation tools or instruments that are used in the appropriate learning outcomes [3]. A good learning evaluation can be carried out through measuring student abilities using test instruments to determine the level of student understanding of the subject matter. Prior to presenting the test instrument, validity and reliability test is first conducted to create a good instrument.

One of the techniques for testing the validity and reliability of test instruments that is popular nowadays is the Rasch Model. This model can classify the calculations between items and students who answer (person) on a distributed map image [4]. Thissen et al. [5] claimed that the Rasch model is part of the item response theory of questions that are given to test takers. This Rasch model focuses on measuring the correlation between the ability of the people answering the questions (person ability) to the difficulty level of the questions (item difficulty) by using a logarithmic function, hence, produce a measurement with the same interval [6]. The Rasch model can recognize high ability in test takers through a greater probability of answering questions correctly than other test takers. Using the Ministep program with the Rasch model can analyze test takers' abilities based on a single parameter. Although using the Rasch model for analysis is quite simple, but the findings are accurate. Rasch investigates the probability of answering correctly. Sumintono & Widhiarso [7] and Linacre [8] stated that The Rasch model has a number of benefits, including the ability to distinguish between respondents who have the same score and identify incorrect responses and missing data scores.

In previous studies, it can be observed that analyzing items using the Rasch model provides a number of benefits [9]. The Rasch model can detect incorrect answers and missing data scores, as well as the fact that ability is not absolutely dependent on correct answers [10]. To determine the quality of an assessment instrument, it has to offer evidences of validity and reliability [11], difficulty level, item discrimination [12] and distractors [13] to show if the instrument can measure what it is expected to measure. Based on these researches, the Rasch Model can be used as a recommended assessment analysis technique in measuring and assessing student learning outcomes to identify students' ability levels. The Rasch model by Winstep software will be used in this study to assess the General Chemistry test instrument's quality.

## **2 Methods**

This research is a descriptive research. The analysis technique in this study is based on the Rasch model to evaluate the test instrument. The goal of this study is to validate the test instrument using the Rasch model in order to analyze the item quality. 29 students from Medan State University's Chemistry Department served as the research samples. The validated test instrument was 25 multiple-choice questions. With the aid of Winstep software 3.73, the Rasch model was employed in the data analysis technique. This model was chosen because it can calculate the likelihood that a student will correctly answer a dichotomous question by comparing their ability to the question's level of difficulty. The validation aspects of the analyzed instruments include the pre-requisite test of item suitability, the level of item difficulty and a student's ability (differential power map), bias test with DIF (differential item function). In this research, the test instrument will be analyzed in terms of the suitability of the items, the reliability, and the measurement information function. However, in this study, the test instrument will be examined for its validity and reliability.

### 3 Result and Discussion

#### 3.1 Validity Test

The implementation of this research activity was initiated by analyzing the items that were used as test instruments in this investigation. The research test instrument needed to be validated to obtain the accuracy between the learning outcome score data and the learning objective by using the Rasch model-based analysis technique. A small-scale experiment was used to validate the test instrument. There are 29 students from Universitas Negeri Medan, Chemistry Department served as the study's sample. There were twenty-five multiple-choice questions on the validated test.. Out of the 25 questions tested for validity, the data collected from the analysis of the validity on Rasch Model-based test instruments can be observed as follows in Table 1.

**Table 1.** Validity analysis result using Rasch Model-based test instruments

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	TOTAL MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXP.	EXACT OBS%	MATCH EXP%	Item
8	2	29	2.70	.74	1.14	.4	3.02	1.9	A-.38	.15	93.1	93.1	8
5	2	29	2.70	.74	1.13	.4	2.87	1.8	B-.33	.15	93.1	93.1	5
15	3	29	2.25	.62	1.19	.5	2.38	1.8	C-.34	.18	89.7	89.6	15
17	5	29	1.63	.50	1.20	.7	2.08	2.1	D-.26	.22	82.8	82.7	17
19	3	29	2.25	.62	1.12	.4	1.79	1.2	E-.15	.18	89.7	89.6	19
9	3	29	2.25	.62	1.07	.3	1.36	.7	F .00	.18	89.7	89.6	9
24	10	29	.64	.41	1.19	1.3	1.18	.9	G .04	.29	58.6	66.7	24
16	9	29	.81	.42	1.09	.6	1.17	.8	H .13	.28	72.4	69.2	16
21	13	29	.16	.39	1.08	.7	1.11	.8	I .20	.32	65.5	64.0	21
12	14	29	.01	.39	1.04	.4	1.03	.3	J .27	.32	58.6	64.0	12
1	24	29	-1.80	.51	1.01	.1	.82	-.3	K .30	.26	82.8	82.7	1
25	18	29	-.62	.40	1.00	.0	.99	.0	L .33	.32	65.5	67.7	25
20	2	29	2.70	.74	.99	.2	.79	.0	M .19	.15	93.1	93.1	20
2	27	29	-2.88	.74	.97	.2	.67	-.2	I .27	.18	93.1	93.1	2
22	23	29	-1.55	.48	.97	.0	.95	.0	k .33	.28	79.3	79.3	22
18	2	29	2.70	.74	.96	.1	.67	-.2	j .25	.15	93.1	93.1	18
14	28	29	-3.63	1.02	.91	.2	.37	-.3	i .35	.13	96.6	96.5	14
3	22	29	-1.34	.45	.90	-.4	.78	-.7	h .46	.29	69.0	75.8	3
6	25	29	-2.08	.55	.87	-.2	.59	-.8	g .48	.24	86.2	86.1	6
23	19	29	-.78	.41	.85	-.9	.87	-.6	f .50	.32	79.3	69.2	23
13	23	29	-1.55	.48	.83	-.6	.67	-.9	e .54	.28	79.3	79.3	13
11	25	29	-2.08	.55	.82	-.4	.59	-.8	d .52	.24	86.2	86.1	11
4	22	29	-1.34	.45	.78	-1.0	.63	-1.3	c .63	.29	75.9	75.8	4
10	10	29	.64	.41	.76	-1.7	.69	-1.7	b .64	.29	79.3	66.7	10
7	24	29	-1.80	.51	.76	-.7	.51	-1.3	a .65	.26	82.8	82.7	7
MEAN	14.3	29.0	.00	.56	.99	.0	1.14	.1			81.4	81.1	
S.D.	9.4	.0	1.94	.15	.14	.6	.71	1.1			10.9	10.5	

Based on the data obtained above, the validity criteria can be analyzed with the Rasch model analysis. For a test instrument to be considered as valid, it must have at least 2 of the 3 criteria that must be fulfilled, which are the following criteria [14]:

- 1)  $0.5 < \text{MNSQ} < 1.5$  : Accepted MNSQ outfit value
- 2)  $-0.2 < \text{ZSTD} < +2.0$  : Accepted ZSTD outfit value
- 3)  $0.4 < \text{Pt Measure Corr} < 0.85$  : Accepted Pt Measure Corr value

According to the results of the analysis by considering these criteria, it was determined that out of 25 items tested for validity, there were 18 items that were stated to be valid and fulfilled at least two valid criteria. The data for the question numbers that were stated to be valid were question numbers 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 16, 18, 20, 21, 22, 23, 24, and question number 25.

### 3.2 Reliability Test

Reliability was also assessed using the Rasch Model with Winstep software. Based on the results of the reliability analysis obtained, it indicates that the quality of the test instrument is categorized as high, which is 0.91 in terms of item reliability. The data from the reliability analysis of the test instrument can be clearly shown in Table 2 below.

**Table 2.** Reliability analysis result of Rasch Model-based test instrument

Item	25 INPUT		25 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	14.3	29.0	.00	.57	.99	.0	1.14	.1
S.D.	9.4	.0	1.94	.16	.14	.6	.71	1.1
REAL RMSE	.59	TRUE SD	1.85	SEPARATION	3.13	Item	RELIABILITY	.91

Based on the data clearly listed in the table, we can see that the item reliability value is 0.91. To determine the level of reliability of the test instrument analyzed using the following criteria [15] on table 3.

**Tabel 3.** The level of reliability test analysis

No	Interval	Criteria
1	< 0.200	Very low
2	0.200 – 0.399	Low
3	0.400 – 0.599	Moderate
4	0.600 – 0.799	High
5	0.800 – 1.000	Very high

It is possible to conclude that the items' reliability level falls into the very high category by analyzing the item reliability criteria..

## 4 Conclusion

1. The results of analysis using the Winstep software based on the Rasch model showed that of the 25 questions tested for validity, 18 questions were declared valid.
2. The results of the analysis using the Winstep software based on the Rasch model obtained the reliability of the items is about 0.91 and categorized as the high category.

**Acknowledgments.** The LPPM at Universitas Negeri Medan provided funding for this research. Thanks to LPPM Universitas Negeri Medan.

## References

- [1] Mardapi, D.: Teknik Penyusunan Instrumen Tes dan Non Tes. Yogyakarta : Mitra Cendikia Offset (2008)
- [2] Purniasari, L., Masykuri, M., & Ariani, S. R. D.: Analisis Butir Soal Ujian Sekolah Mata Pelajaran Kimia SMA N 1 Kutowinangun Tahun Pelajaran 2019/2022 Menggunakan Model Iteman dan Rasch. *Jurnal Pendidikan Kimia*, 10(2), 205–214 (2021)
- [3] Rusiyah, Eraku, S. S., & Supadmi, S.: Analisis Soal Ujian Akhir Semester Mata Pelajaran Geografi Dengan Menggunakan Pemodelan Rasch. *JURNAL SWARNABHUMI: Jurnal Geografi Dan Pembelajaran Geografi*, 5(1), 11 (2020)
- [4] Rozeha, A. R., Azami, Z., & Saidfudin, M.: Application of Rasch Measurement in Evaluation of Learning Outcomes: A Case Study in Electrical Engineering. *Regional Conference on Engineering Mathematics, Mechanics, Manufacturing & Architecture* (2007)
- [5] Thissen, D., Nelson, L., Rosa, K.: Item Response Theory for Items Scored in More than Two Categories dalam D. Thissen & H. Wainer, *Test Scoring* (pp. 141-184). New Jersey: Lawrence Erlbaum Associates Publishers (2001)
- [6] Olsen, L. W.: Essay son Georg Rasch and his contribution statistics. Un published PhD the sisat Institute Of Economics University of Copenhagen (2003)
- [7] Sumintono, B., & Widhiarso, W.: Aplikasi Pemodelan Rasch Pada Assessment Pendidikan. *Trim Komunikata* (2015)
- [8] Linacre, J.M.: A User's Guide to WINSTEPS MINISTEP Rasch-Model Computer Programs, IL: Winsteps.com, Chicago (2016)
- [9] Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa, I., & Nurhaswinda, N.: Analysis of RASCH model for the validation of chemistry national exam instruments. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, 9(3), 329-345 (2021)
- [10] Eliza, W., & Yusmaita, E.: Pengembangan Butir Soal Literasi Kimia pada Materi Sistem Koloid Kelas XI IPA SMA/MA. *Jurnal Eksakta Pendidikan (JEP)*, 5(2), 197-204 (2021)
- [11] Sofia, Permasari, A., Sholihin, H., & Supriyanti, F. M. T.: Konten Dan Konstruksi Asesmen Keterampilan Berpikir Evaluatif Pada Tema Karbohidrat Dalam Kimia Pangan. *Orbital: Jurnal Pendidikan Kimia*, 4(1), 88–98 (2020)
- [12] Son, A. L.: Instrumentasi Kemampuan Pemecahan Masalah Matematis: Analisis Reliabilitas, Validitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Gema Wiralodra*, 10(1), 41–52 (2019)
- [13] Damanik, F. C., & Irfandi.: Pengembangan Instrumen Tes Berbasis Higher Order Thinking Skill (HOTS) Bentuk Pilihan Ganda Berdasarkan Taksonomi Bloom Revisi Pada Materi Gelombang Bunyi. *Jurnal Inovasi Pembelajaran Fisika*, 1 0(1) (2021).

- [14] Tarigan, E. F., Nilmarito, S., Islamiyah, K., Darmana, A., & Suyanti, R. D.: Analisis Instrumen Tes Menggunakan Rasch Model dan Software SPSS 22.0. *Jurnal Inovasi Pendidikan Kimia*, 16(2), 92-96 (2022)
- [15] <https://slideplayer.info/amp/2988918/>