

Rasch Modeling: A Multiple Choice Chemistry Test

Atiek Winarti¹, Al Mubarak²

^{1,2}Faculty of Teacher Training Education, Universitas Lambung Mangkurat, Indonesia

DOI: 10.23917/ijolae.v2i1.8985

Received: October 6th, 2019. Revised: October 23rd, 2019. Accepted: October 25th, 2019.

Available Online: October 26th, 2019. Published Regularly: January 1st, 2020.

Abstract

The study aimed to reveal the difficulty level of items and the suitability of items of Chemistry test with the Rasch model. In addition to detecting this item quality, the Rasch model shows the student's answer pattern as well, so that the assessment can imply the quality of the instrument as an assessment of chemical learning. As many as 20 numbers of multiple-choice questions in chemical bonding material were analyzed by using WINSTEPS 3.73. The samples consisted of 200 senior high school students in Banjarmasin Indonesia. The results revealed that the average item measure was 0.00 with items (Measure Order = 4.64) which has the highest difficulty level. The Q10 was the item that has a level of conformity with the model, and outliers or misfit in Rasch were MNSQ=+0.97, ZSTD=-0.2, Pt Mean Corr=+0.58. In other words, assessment of learning with test techniques such as multiple choice based on Rasch model analysis was an effective way for teachers to review the progress of students in the learning process, guidelines for designing chemical learning strategies, and identifying students' understanding of chemical material.

Keywords: rasch model, multiple choices, chemical bonding

Corresponding Author:

Atiek Winarti, Faculty of Teacher Training Education, Universitas Lambung Mangkurat, Indonesia

e-mail: atiekwin_kimia@ulm.ac.id

1. Introduction

Learning assessment is an important point that must be conducted by the teacher (Potgieter & Davidowitz, 2011). The assessment of learning provided contains a lot of essential information such as evaluating student learning progress, the extent of students' cognitive depth on learning that has been passed, and the accuracy of test instruments in measuring their mental models (Potgieter & Davidowitz, 2011; Brannon et al, 2018). The learning process that is assessed by the instructor without using assessment will be difficult to know cognitive development, behavior, depth of understanding, and the impact of the designed teaching (Potgieter & Davidowitz, 2011).

During the learning process, the instructor will not only transform knowledge in the classroom but also provide other treatments

to develop students' potential such as assignments, practice cognitive skills, formative or summative tests, and direct communication between the teacher and students scientifically (Sprague et. Al, 2018). Indirectly, that learning assessment is an illustration of how the teacher reflects on the learning process that has been experienced (Izci et al., 2018). Associated with the teaching of chemistry, that the main target of learning is not only to bring students at a high cognitive level but also how students interpret and internalize experiences that have been experienced while learning to become valuable individuals (Hindal, 2013).

A written test as an assessment technique is often used to review the progress of student learning and the effectiveness of learning undertaken (Herrmann-Abell & DeBoer, 2011). Written tests that can be used are multiple choice with the aim of analyzing

the cognitive processes of students, emphasizing whether students experience misconceptions or not on chemical material, and evaluating learning concepts, so that assessment results are used as references in updating the learning process, and identifying the nature of the items distributed (Brandriet & Bretz, 2014; Cheung, 2011; Treagust, et al, 2011; Milenković, et al, 2016; Yasin, et al, 2015). Experts also emphasized that the use of a distractor for multiple choice instruments could be used as an advantage in making multiple choice items (Herrmann-Abell & DeBoer, 2011; Villafaña, 2011). Multiple choice questions with distractors can diagnose students' ability to understand material and this becomes a strategy in preventing potential student misconceptions (Herrmann-Abell & DeBoer, 2011). In addition, the instructor also made the results of cognitive chemical analysis in this format as a guide in evaluating the learning process (Rauch & Hartig, 2010).

Measurements with the multiple choice question format are closely related to the ability of the instrument to measure students' cognitive structure and item quality (de la Torre, 2009; Zamri, 2015). That is, the aspects of validity and reliability are part of the interpretation of data, especially the quality of the items (Zamri, 2015; Brandriet et.al, 2015). The Rasch model is a data analysis technique where it is very effective, precise, and systematic in justifying items with a logarithmic approach (Yasin et. al, 2015; Park et al. 2017; Lee et al., 2011). The concept of the Rasch model is not just a technique that shows the level of difficulty and feasibility of the question assessment instrument, but it is also able to show the pattern of student answers in responding to the problem (Sumintono, 2018; Chiang, 2015). In other words, the use of the Rasch model confirms that each student has the same opportunity to answer questions correctly, not just about items (Park et al. 2017; Chan et al, 2014). Rasch not only assesses the ability of students through the distribution of data, but also shows the level of difficulty of the problem, the suitability of

the items with the sample used, and the symptoms of misconception, so that the Rasch model is an appropriate analysis technique used to identify the things mentioned (Zamri, 2015). The conclusion is that the Rasch model provides a more comprehensive and concrete picture in the measurement aspects of the test because the Rasch model involves two basic aspects as parameters namely (a) students' abilities, and (b) the difficulty level of the question or ability of the question (Zamri, 2015; Amin et al, 2012; Runnels, 2012).

2. Method

The quantitative approach was used. It was used because the instructor needed to ascertain and confirm what and how the assessment process is carried out so that it required a quantitative and in-depth interpretation of the research conducted (Potgieter & Davidowitz, 2011). The selected samples are 200 students of chemistry education department from semester 1 to semester 5. These samples were considered important to identify their potential as prospective teachers.

The assessment instrument of Chemical bonding material consisted of 20 number of multiple choice questions were analyzed by using WINSTEPS 3.73 Rasch model. The questions analyzed would represent the level of students' understanding of the material and also determined the measurement of the questions. Rasch modeling with multiple choice formats uses and combines an algorithm that states the results of probabilistic expectations of items "i" and respondents "n", which are mathematically expressed as (Chan, 2014; Runnels, 2012):

$$P_{ni} (X_{ni} = 1 | b_n, d_i) = \frac{e^{(b_n - d_i)}}{1 + e^{(b_n - d_i)}}$$

Where $P_{ni} (X_{ni} = 1 | b_n, d_i)$ is the probability of respondent n in item i to produce a correct answer ($x = 1$); with the respondent's ability, β_n and the difficulty level of the item δ_i . The equation above by Rasch can be further simplified by entering the logarithmic function and making it:

$$\text{Log } P_{ni} (X_{ni} = 1 | \beta_n, \delta_i) = \beta_n - \delta_i$$

So the probability will be a success that can be written as:

Probability to succeed = respondent's ability - item difficulty level

The Rasch model emphasizes that each student has the same opportunity to answer questions correctly and at the same time the problem has different levels of difficulty. This is termed Rasch as a person logit and logit item.

Person logit = $\ln[p/(1-p)]$

Item logit = $\ln [p\text{-value}] = \ln [p\text{-value}/(1-p\text{-value})]$

3. Result and Discussion

a. Item Measure

The item measure in Rasch is an analysis of the difficulty level of the item. The rightmost column (item) is the code of the 20 items distributed. This table provides a lot of information about the items distributed such as the "measure" column, where the column visualizes the difficulty level of each item with the term "logit value". From top to bottom is the range from the difficulty level of the highest item to the lowest level. It means item with code Q15 is the item with the highest difficulty level (logit = 4.64) and Q7 is the item with the lowest difficulty level (-3.45 logit). Evidence that Q15 is the most difficult question, namely in the "total score" column, where Q15 has a total score of 7. This score indicates that overall only 7 samples answered questions Q15 correctly. Conversely, item Q7 has a total score of 192, meaning that 192 samples are able to answer Q7 questions correctly. In other words, Q7 is a question that is easily answered by students.

Then, the table can also explain the specifications of the items in another perspective because Rasch produces the same scale distance. If you pay attention to the "logit value" of the question, they

(questions) have different logit values between one another. For example, the Q19 (logit value = 1.93) when compared to the Q17 (logit value = .53) and the Q5 (logit value = 14), it can be said that the Q19 has a difficulty level 3 times compared to the Q17 and more than 10 times the Q5. Besides the distance of logit values differed significantly, the total score value also looks much different, where only 57 samples answered correctly for item Q19, 107 samples answered correctly for item Q17, and Q5 as many as 121 samples answered correctly.

Overall, the item measure table has provided concrete information about the condition of each item so that this can be the evaluation for the teacher and the students. Teacher can use this data as a guideline and reference in making the appropriate learning assessment (written test assessment), knowing the characters of each question, analyzing the extent of students' understanding of the chemical bonding material, and how each item is assessed as a learning assessment.

b. Item Fit

The "misfit order" in Rasch is an analysis of the suitability of the item. If we previously discussed the difficulty level of the item, then "order misfit" represents the item's suitability level. Item fit (misfit order) explains whether an item is functioning normally or not in making measurements. If the detected items are not fit, then the question indicates that there is a misconception between students and the items they are working on (Herrmann-Abell & DeBoer, 2011). The indicated items are not fit, it needs more in-depth discussion. It means the teacher needs to further analyze the problem structure and student answer patterns so that they can find a point of the problem. Information based on this table is very valuable for teachers to be a reference for improving the quality of teaching. In addition, this information is also able to prevent and deal with attacks of misconceptions that will arise next.

Table 1. Item measure

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL		INFIT		OUTFIT		PT-MEASURE		EXACT OBS%	MATCH EXP%	Item
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
15	7	200	4.64	.39	1.14	.5	3.73	3.0	-.07	.17	96.5	96.5	Q15	
16	41	200	2.46	.19	1.59	5.1	2.93	4.6	-.12	.37	72.5	80.5	Q16	
19	57	200	1.93	.18	1.08	1.1	1.24	1.1	.35	.42	72.0	75.7	Q19	
14	66	200	1.66	.17	.90	-1.4	.73	-1.6	.53	.45	74.5	73.8	Q14	
11	98	200	.78	.16	.91	-1.4	.83	-1.4	.57	.50	75.5	72.1	Q11	
3	105	200	.59	.17	1.09	1.3	1.05	.5	.46	.51	69.0	72.6	Q3	
17	107	200	.53	.17	1.29	3.7	1.33	2.5	.34	.51	60.0	72.7	Q17	
13	112	200	.39	.17	.75	-3.7	.71	-2.6	.67	.52	86.5	73.4	Q13	
12	120	200	.17	.17	.86	-1.8	.73	-2.3	.62	.52	76.0	74.4	Q12	
5	121	200	.14	.17	.90	-1.3	.90	-.7	.58	.52	82.5	74.5	Q5	
6	129	200	-.10	.17	.99	-.1	.88	-.9	.54	.52	74.5	75.8	Q6	
4	132	200	-.19	.18	.70	-3.8	.56	-3.7	.71	.52	83.0	76.5	Q4	
10	134	200	-.25	.18	.88	-1.3	.97	-.2	.58	.52	79.0	76.9	Q10	
9	146	200	-.65	.19	1.16	1.6	1.28	1.5	.41	.51	79.0	79.9	Q9	
2	153	200	-.91	.20	.83	-1.6	.59	-2.3	.63	.51	84.5	82.0	Q2	
20	164	200	-1.37	.21	.64	-3.2	.41	-2.8	.69	.48	90.5	85.5	Q20	
8	168	200	-1.56	.22	1.08	.6	.86	-.4	.45	.47	84.5	86.9	Q8	
1	181	200	-2.34	.27	1.21	1.1	1.00	-.2	.32	.41	88.5	91.1	Q1	
18	183	200	-2.49	.28	.87	-.6	.81	-.3	.44	.39	94.5	91.8	Q18	
7	192	200	-3.45	.39	1.11	.5	.91	.1	.22	.29	96.0	96.0	Q7	
MEAN	120.8	200.0	.00	.21	1.00	-.3	1.12	-.3			80.9	80.4		
S.D.	48.0	.0	1.80	.07	.22	2.2	.78	2.0			9.3	7.9		

Table 2. Item Fit/Misfit Order

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL		INFIT		OUTFIT		PT-MEASURE		EXACT OBS%	MATCH EXP%	Item
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
15	7	200	4.64	.39	1.14	.5	3.73	3.0	A-.07	.17	96.5	96.5	Q15	
16	41	200	2.46	.19	1.59	5.1	2.93	4.6	B-.12	.37	72.5	80.5	Q16	
17	107	200	.53	.17	1.29	3.7	1.33	2.5	C .34	.51	60.0	72.7	Q17	
9	146	200	-.65	.19	1.16	1.6	1.28	1.5	D .41	.51	79.0	79.9	Q9	
19	57	200	1.93	.18	1.08	1.1	1.24	1.1	E .35	.42	72.0	75.7	Q19	
1	181	200	-2.34	.27	1.21	1.1	1.00	-.2	F .32	.41	88.5	91.1	Q1	
7	192	200	-3.45	.39	1.11	.5	.91	.1	G .22	.29	96.0	96.0	Q7	
3	105	200	.59	.17	1.09	1.3	1.05	.5	H .46	.51	69.0	72.6	Q3	
8	168	200	-1.56	.22	1.08	.6	.86	-.4	I .45	.47	84.5	86.9	Q8	
6	129	200	-.10	.17	.99	-.1	.88	-.9	J .54	.52	74.5	75.8	Q6	
10	134	200	-.25	.18	.88	-1.3	.97	-.2	J .58	.52	79.0	76.9	Q10	
11	98	200	.78	.16	.91	-1.4	.83	-1.4	i .57	.50	75.5	72.1	Q11	
5	121	200	.14	.17	.90	-1.3	.90	-.7	h .58	.52	82.5	74.5	Q5	
14	66	200	1.66	.17	.90	-1.4	.73	-1.6	g .53	.45	74.5	73.8	Q14	
18	183	200	-2.49	.28	.87	-.6	.81	-.3	f .44	.39	94.5	91.8	Q18	
12	120	200	.17	.17	.86	-1.8	.73	-2.3	e .62	.52	76.0	74.4	Q12	
2	153	200	-.91	.20	.83	-1.6	.59	-2.3	d .63	.51	84.5	82.0	Q2	
13	112	200	.39	.17	.75	-3.7	.71	-2.6	c .67	.52	86.5	73.4	Q13	
4	132	200	-.19	.18	.70	-3.8	.56	-3.7	b .71	.52	83.0	76.5	Q4	
20	164	200	-1.37	.21	.64	-3.2	.41	-2.8	a .69	.48	90.5	85.5	Q20	
MEAN	120.8	200.0	.00	.21	1.00	-.3	1.12	-.3			80.9	80.4		
S.D.	48.0	.0	1.80	.07	.22	2.2	.78	2.0			9.3	7.9		

Boone et al (2014) explained that checking the suitability of the items can be identified based on the Mean Square Outfit (MNSQ) value with a range of values of $0.5 < \text{MNSQ} < 1.5$, Z-standard (ZSTD) outfit values for the range $-2.0 < \text{ZSTD} < +2.0$, and $0.4 < \text{Pt Mean Corr} < 0.85$ for the value of Point Measure Correlation outfit (Pt Mean Corr). The value of ZSTD is strongly influenced by the number of samples used. When the sample size is large, then the ZSTD outfit value will always be above 3 so that the ZSTD value by experts does not recommend this value when the sample size in calibration is large ($N = 500$).

Table 1 shows that Q15 item is on the first order of all items analyzed. This indicates that based on the "item measure" table, the Q15 is the item that has the highest difficulty level among the other items (logit value = + 4.64). Then, fit items are the basis for assessing whether or not the item is maintained as a measuring instrument.

This is indicated by their outfit value, where the Q15 has an MNSQ outfit value of 3.73, ZSTD outfit value is 3.0, and -0.07 for the Pt Mean Corr outfit value. If we look back at the standard criteria for item suitability, the outfit value of Q15 is an item that is not fit. It means the Q15 is not statistically recommended to be used as a measurement tool because it is not representative in measuring students' understanding of chemical bonding material. The reason that Q15 is the hardest question is because of from 200 samples, only 7 samples answered the Q15 correctly (see "total score" in the item fit / item measure table).

Another proof to support the argument above is through the suitability of the items that can be seen from the graph, namely the "expected ICC score" chart. Look at the graph of the specific item's suitability for items for the Q15, where there is a response beyond the Rasch point threshold.

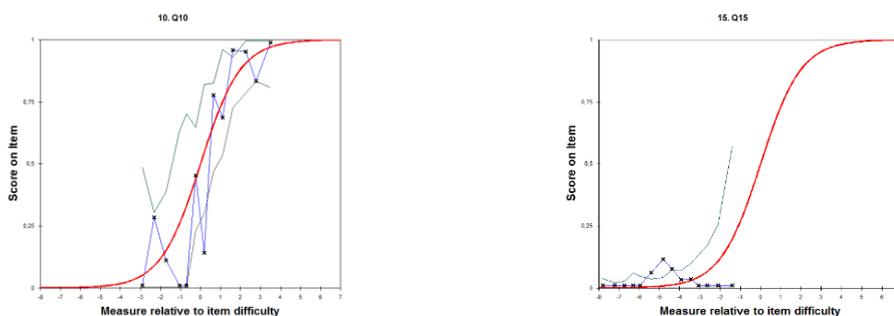


Figure 1. Graph of item adjustment level for Q15

Based of this finding, it can be concluded that Q15 is not recommended to be used as a measuring instrument. Whereas, number 15 is content as a basic material in understanding the concept of chemical bonds. The following is question number 15.

The amount of covalent bonds in H_2SO_4 molecules is ... (answer B).

- a. 1 d. 0 (none)
 b. 4 e. 3
 c. 2

The basic theory in answering the above questions is, first: students must have the knowledge aspect at level C1 (remembering), where they have to recall how to write shell and subshell electrons configuration for each chemical element. Then, they have to write the configuration precisely and based on scientific principles of the electron configuration theory in question. The second, they are able to analyze the symptoms and nature of a molecule based on configuration, this will stimulate cognitively and construct their mind so that the role of valence electrons is very much needed. In addition, they must know the concept of "formal charge" and how to implement the concept. If they know and understand the points mentioned, students are able to analyze and identify the tendency of each chemical element/molecule to bond.

Based on the item structure, actually the Q15 can be considered as the easiest item for university students because they have a more mature mental model of thinking. But, only the small number of students who answer

this item corectly. Student's cognitive abilities in analyzing the Q15 question are not fully optimal even though these questions are considered as the most basic scientific questions to understand in the material of chemical bonds. Representative interpretation especially sub-microscopic and symbolic levels (Eliyawati et al, 2018; Serobatse, et al, 2014) can help students answer Q15 correctly. When viewed in a matter structure, Q15 is an easy question to be answered by college students where they already have a more mature mental model of thinking. That is, their cognitive abilities in analyzing Q15 questions are not fully optimal even though these questions are considered as the most basic scientific questions to understand in the material of chemical bonds.

Another analysis is the Q10, where this item is in the fit criteria, misfit criteria, which has MNSQ outfit values of 0.97, ZSTD -0.2, and Pt Mean Corr 0.58. The Q10 can be used as a measurement tool in analyzing students' cognitive. Overall, Rasch-based statistical results, measuring student understanding, have provided a lot of information.

Consider the following table!

No	Compound Pairs
1	NH_4^+ and NH_3BF_3
2	H_2O and HCl
3	HF and H_2O
4	NH_3 and HBr
5	NH_3 and H_2S

The group of compounds that have coordination covalent bonds is

- | | |
|------|-------------|
| a. 1 | d. 0 (None) |
| b. 4 | e. 3 |
| c. 2 | |

Overall the statistical results shown that the Rasch model analysis provide a lot of information about the feasibility of an item in measuring the level of student understanding. This is because, if we examine more deeply the results of the analysis of the Rasch model, then we not only find the size of the instruments made but also the patterns of student answers. As a result, teachers obtain a lot of information as a basis for improving the quality of teaching and as an evaluation of chemistry learning. This is reinforced by research (Cheung, 2011) that it is necessary to carry out a process of diagnosis of students and the learning process carried out as a strategy to design effective teaching and improve the quality of students' understanding in learning chemistry. In addition, (Zamri bin Khairani & Bin Abd. Razak, 2015) added that the Rasch model is able to provide more interpretation of the results of the statistical data presented so that teachers are easier in the diagnosis process.

4. Conclusion

Overall it can be concluded that the distributed dichotomy items are worth to be used as the assessment tool of cognitive ability by viewing the item fit value. The item fit data showed that 90% of the questions are worth to be used as the assessment tool. From 20 numbers of analyzed questions the item of Q15, Q16, and Q17 are not suitable with Rasch model criteria. Even though ZSTD value is affected by the number of sample, MNSQ and Pt Mean Corr values become the other references in assessing the quality of the questions.

5. References

Brandriet, A. R., & Bretz, S. L. (2014). The

Development of the Redox Concept Inventory as a Measure of Students' Symbolic and Particulate Redox Understandings and Confidence. *Journal of Chemical Education*, 91, 1132–1144.

Brandriet, A. R., Xu, X., Bretz, S. L., & Lewis, J. E. (2011). Diagnosing changes in attitude in first-year college chemistry students with a shortened version of Bauer's semantic differential. *Chemistry Education Research and Practice*, 12(2), 271–278. <https://doi.org/10.1039/c1rp90032c>

Brannon, R. M., Rusilowati, A., & Nugroho, S. E. (2018). Design of Chemical Literacy Assessment by Using Model of Educational Reconstruction (MER) on Solubility Topic Design of Chemical Literacy Assessment by Using Model of Educational Reconstruction (MER) on Solubility Topic. *IOP Conference Series: Material Science and Engineering*. <https://doi.org/10.1088/1757-899X/335/1/012106>

Bruce, M. R. M., Bruce, A. E., Avargil, S., Amar, G., Wemyss, T. M., & Flood, V. J. (2016). Polymers and Cross-Linking: A CORE Experiment To Help Students Think on the Submicroscopic Level. *Journal of Chemical Education*, 93(9), 1599–1605. <https://doi.org/10.1021/acs.jchemed.6b00010>

Chan, S. W., Ismail, Z., & Sumintono, B. (2014). A Rasch Model Analysis on Secondary Students' Statistical Reasoning Ability in Descriptive Statistics. *Procedia - Social and Behavioral Sciences*, 129, 133–139. <https://doi.org/10.1016/j.sbspro.2014.03.658>

Cheung, D. (2011). Using diagnostic assessment to help teachers understand the chemistry of the lead-acid battery.

- Chemistry Education Research and Practice*, 12(2), 228–237. <https://doi.org/10.1039/c1rp90028e>
- Chiang, W.-W. (2015). Ninth Grade Student' Self-assessment in Science: A Rasch Analysis Approach. *Procedia - Social and Behavioral Sciences*, 176, 200–210. <https://doi.org/10.1016/j.sbspro.2015.01.462>
- Chow, J., Tse, A., & Armatas, C. (2018). Comparing trained and untrained teachers on their use of LMS tools using the Rasch analysis. *Computers and Education*, 123, 124–137. <https://doi.org/10.1016/j.compedu.2018.04.009>
- Cloonan, C. A., & Hutchinson, J. S. (2011). A chemistry concept reasoning test. *Chemistry Education Research and Practice*, 12(2), 205–209. <https://doi.org/10.1039/c1rp90025k>
- Eliyawati, Rohman, I., & Kadarohman, A. (2018). The effect of learning multimedia on students' understanding of macroscopic, sub-microscopic, and symbolic levels in electrolyte and nonelectrolyte. *Journal of Physics: Conference Series*, 1013(1). <https://doi.org/10.1088/1742-6596/1013/1/012002>
- Erdogan, T., & Senemoglu, N. (2014). Problem-Based Learning in Teacher Education: Its Promises and Challenges. *Procedia - Social and Behavioral Sciences*, 116, 459–463. <https://doi.org/10.1016/j.sbspro.2014.01.240>
- Flood, V. J., Bruce, M. R. M., & Wittmann, M. C. (2015). Paying Attention to Gesture when Students Talk Chemistry: Interactional Resources for Responsive Teaching. *Journal of Chemical Education*, 92(1), 11–22.
- Henning, G. (1989). Does the Rasch Model Really Work for Multiple-Choice Items? Take Another Look: A Response to Divgi. *Journal of Educational Measurement*, 26(1), 91–97. Retrieved from <http://links.jstor.org/sici?sici=0022-0655%28198921%2926%3A1%3C91%3ADTRMRW%3E2.0.CO%3B2-6>
- Herrington, D. G., & Sweeder, R. D. (2018). Outside of the Classroom. *Journal of Chemical Education*, 95(12), 2148–2154. <https://doi.org/10.1021/acs.jchemed.8b00361>
- 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. *Chemistry Education Research and Practice*, 12(2), 184–192. <https://doi.org/10.1039/c1rp90023d>
- Kiliç, D., & Sağlam, N. (2009). Development of a two-tier diagnostic test concerning genetics concepts: the study of validity and reliability. *Procedia - Social and Behavioral Sciences*, 1(1), 2685–2686. <https://doi.org/10.1016/j.sbspro.2009.01.474>
- Lee, H. S., Liu, O. L., & Linn, M. C. (2011). Validating measurement of knowledge integration in science using multiple-choice and explanation items. *Applied Measurement in Education*, 24(2), 115–136. <https://doi.org/10.1080/08957347.2011.554604>
- Mamat, M. N., Maidin, P., & Mokhtar, F. (2014). Simplified Reliable Procedure for Producing Accurate Student's Ability Grade Using Rasch Model. *Procedia - Social and Behavioral Sciences*, 112(Icepsy 2013), 1077–1082. <https://doi.org/10.1016/j.sbspro.2014.01>

- .1272
- Milenkovic, D. Dusica., Segedinac, D. Mirjana., and Hrin, N. T. (2014). Increasing High School Students' Chemistry Performance and Reducing Cognitive Load through an Instructional Strategy Based on the Interaction of Multiple Levels of Knowledge Representation. *Journal of Chemical Education*, 91(9), 1409–1416.
- Milenković, D. D., Hrin, T. N., Segedinac, M. D., & Horvat, S. (2016). Identification of Misconceptions through Multiple Choice Tasks at Municipal Chemistry Competition Test Identification of Misconceptions through Multiple Choice Tasks at Municipal Chemistry Competition Test. *Journal of Subject Didactics*, 1(November), 3–12. <https://doi.org/10.5281/zenodo.55468>
- Naiker, M., & Wakeling, L. (2015). Evaluation of group based inquiry oriented learning in undergraduate chemistry practicals. *International Journal of Innovation in Science and Mathematics Education*, 23(5), 1–17.
- Pappa, E. T., & Tsapralis, G. (2011). Evaluation of questions in general chemistry textbooks according to the form of the questions and the question-answer relationship (QAR): The case of intra- and intermolecular chemical bonding. *Chemistry Education Research and Practice*, 12(2), 262–270. <https://doi.org/10.1039/c1rp90031e>
- Park, M., Liu, X., & Waight, N. (2017). Development of the Connected Chemistry as Formative Assessment Pedagogy for High School Chemistry Teaching. *Journal of Chemical Education*, 94(3), 273–281. <https://doi.org/10.1021/acs.jchemed.6b00299>
- Potgieter, M., & Davidowitz, B. (2011). Preparedness for tertiary chemistry: Multiple applications of the Chemistry Competence Test for diagnostic and prediction purposes. *Chemistry Education Research and Practice*, 12(2), 193–204. <https://doi.org/10.1039/c1rp90024b>
- Runnels, J. (2012). Using the Rash model to validate a multiple choice English achievement test. *International Journal of Language Studies*, 6(4), 141–155.
- Serobatse, B. M., Selvaratnam, M., & Drummond, H. P. (2014). Students' conceptions about the sub-microscopic approach to explanations in chemistry throughout their BSc degree course: research article. *South African Journal of Chemistry*, 67, 40–44. Retrieved from http://reference.sabinet.co.za/webx/access/electronic_journals/chem/chem_v67_a7.pdf
- Sprague, E., Siegert, R. J., Medvedev, O., & Roberts, M. H. (2018). Rasch Analysis of the Edmonton Symptom Assessment System. *Journal of Pain and Symptom Management*, 55(5), 1356–1363. <https://doi.org/10.1016/j.jpainsymman.2018.01.016>
- Sukor, N. S., Osman, K., & Abdullah, M. (2010). Students' achievement of Malaysian 21st Century Skills in Chemistry. *Procedia - Social and Behavioral Sciences*, 9, 1256–1260. <https://doi.org/10.1016/j.sbspro.2010.12.316>
- Sumintono, B. (2018). Rasch Model Measurements as Tools in Assesment for Learning, (October 2017). <https://doi.org/10.2991/icei-17.2018.11>
- Torre, J. de la. (2009). A Cognitive Diagnosis Model for Cognitively Based Multiple-Choice Options. *Applied Psychological Measurement*, 33(3), 163–183. <https://doi.org/10.1177/0146621608320523>

- Üce, M., & Ceyhan, İ. (2019). Misconception in Chemistry Education and Practices to Eliminate Them: Literature Analysis. *Journal of Education and Training Studies*, 7(3), 202–208. <https://doi.org/10.11114/jets.v7i3.3990>
- Velychko, O., & Gordiyenko, T. (2018). A comparative analysis of the assessment results of the competence of technical experts by methods of analytic hierarchy process and with using the Rasch model. *Eastern-European Journal of Enterprise Technologies*, 3(3 (93)), 14–21. <https://doi.org/10.15587/1729-4061.2018.131459>
- Versprille, A., Zabih, A., Holme, T. A., Mckenzie, L., Maha, P., Martin, B., & Towns, M. (2017). Assessing Student Knowledge of Chemistry and Climate Science Concepts Associated with Climate Change: Resources To Inform Teaching and Learning. *Journal of Chemical Ed*, 94(4), 407–417. <https://doi.org/10.1021/acs.jchemed.6b00759>
- Villafañe, S. M., Loertscher, J., Minderhout, V., & Lewis, J. E. (2011). Uncovering students' incorrect ideas about foundational concepts for biochemistry. *Chemistry Education Research and Practice*, 12(2), 210–218. <https://doi.org/10.1039/c1rp90026a>
- Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using Rasch Measurement To Develop a Computer Modeling-Based Instrument To Assess Students' Conceptual Understanding of Matter. *Journal of Chemical Education*, 89, 335–345.
- Wiliam, D. (2011). Studies in Educational Evaluation What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Yasin, R. M., Rus, R. C., Ahmad, A., Rahim, M. B., & Yunus, F. A. N. (2015). Validity and Reliability Learning Transfer Item Using Rasch Measurement Model. *Procedia - Social and Behavioral Sciences*, 204(November 2014), 212–217. <https://doi.org/10.1016/j.sbspro.2015.08.143>
- Zamri bin Khairani, A., & Bin Abd. Razak, N. (2015). Modeling a Multiple Choice Mathematics Test with the Rasch Model. *Indian Journal of Science and Technology*, 8(12). <https://doi.org/10.17485/ijst/2015/v8i12/70650>