

# The Use of The Rasch Model to Develop Students' Conception of Chemistry Learning Instruments During the Covid-19 Pandemic

Ifan Rivaldo<sup>1\*</sup>, Hari Sutrisno<sup>2</sup>, Alusti Cundo Manik<sup>3</sup> 

<sup>1,2,3</sup> Program Studi Pendidikan Kimia, Universitas Negeri Yogyakarta, Indonesia

## ARTICLE INFO

### Article history:

Received January 16, 2022

Revised January 25, 2022

Accepted March 22, 2022

Available online April 25, 2022

### Kata Kunci:

Konsep Belajar, Pembelajaran Kimia, Model Rasch, Pandemi Covid-19

### Keywords:

Conception of Learning, Chemistry Learning, Rasch Model, Covid-19 Pandemic



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright © 2022 by Author. Published by Universitas Pendidikan Ganesha.

## ABSTRAK

Perbedaan konsepsi belajar tergantung pada pengalaman belajar kimia yang dirasakan peserta didik. Perbedaan ini semakin tampak ketika peserta didik harus menjalani pembelajaran jarak jauh dikarenakan pandemi. Untuk mengukur perbedaan tersebut, instrumen konsepsi belajar kimia yang terdiri dari 24 item dengan skala likert dikembangkan dan divalidasi oleh para ahli secara konstruk serta secara empiris dengan menerapkan analisis model Rasch menggunakan software WINSTEPS® (versi 3.73). Tujuan penelitian ini yaitu mengembangkan instrumen konsepsi belajar yang spesifik mengukur konsepsi belajar kimia. Jenis penelitian ini yaitu pengembangan. Populasi penelitian ini sejumlah 247 siswa sekolah menengah atas. Metode yang digunakan untuk mengumpulkan data yaitu survei. Instrumen yang digunakan untuk mengumpulkan data yaitu kuesioner. Kualitas psikometrik kuesioner yang dianalisis meliputi reliabilitas, statistik kesesuaian item, penggunaan skala penilaian, dan bias item terhadap gender (Differential Item Functioning (DIF)). Teknik yang digunakan untuk menganalisis data yaitu analisis deskriptif kualitatif dan kuantitatif. Hasil penelitian menunjukkan bahwa model Rasch cocok dengan data pengukuran meliputi reliabilitas person dan item (>0,8). Sembilan belas item memenuhi nilai penerimaan untuk statistik kesesuaian item. Selain itu, kategori pada skala berfungsi dengan baik dan bebas dari ketidakteraturan dari nilai Andrich-Threshold. 2 dari 4 item bias gender kemudian dipertahankan dengan merevisi kuesioner. Disimpulkan bahwa instrumen valid dan dapat digunakan untuk mengukur konsepsi belajar kimia.

## ABSTRACT

Different conceptions of learning depend on the chemistry learning experience felt by students. This difference is increasingly visible when students undergo distance learning due to the pandemic. The conceptualization of a chemistry learning instrument consisting of 24 items with a Likert scale was developed and validated by experts in a constructivist and empirical manner by applying the Rasch model analysis using the WINSTEPS® software (version 3.73). This research aims to develop a learning conception instrument that explicitly measures the conception of chemistry learning. This type of research is development. The population of this study was 247 high school students. The method used to collect data is a survey. The instrument used to collect data is a questionnaire. The psychometric quality of the questionnaires analyzed included reliability, item suitability statistics, use of rating scales, and item bias towards gender (Differential Item Functioning (DIF)). The technique used to analyze the data is descriptive qualitative, and quantitative analysis. The results showed that the Rasch model fits the measurement data, including the person and item reliability (> 0.8). Nineteen items met the acceptance scores for the item suitability statistic. In addition, the categories on the scale function well and are free from irregularities from the Andrich-Threshold value. 2 of the 4 gender-biased items were then maintained by revising the questionnaire. It was concluded that the instrument was valid and could be used to measure the concept of chemistry learning.

## 1. INTRODUCTION

Learning during the pandemic has become a challenge for various parties, including the school's organizers, the government, parents, and students. Schools must continue carrying out distance learning with adequate use of technology by teachers and students (Fadhilaturrahm et al., 2021; Wijayanti & Fauziah, 2020). The government has made good efforts to provide infrastructure to support distance learning (Churiyah et al., 2020; Fadhilaturrahm et al., 2021; Sa'diyah, 2021). The Ministry of Education and Culture provides 23 online learning pages such as Rumah Belajar, TV Edukasi, etc. However, KPAI (Komisi Perlindungan Anak Indonesia) reports that many students experience stress during distance learning due to too many learning assignments given by the teacher. This condition provides an overview of distance learning in all subjects, including chemistry. Chemistry is considered one of the difficult and abstract subjects that creates fear for students in learning (Refriwati, 2015; Zammiluni et al., 2018). It is exacerbated by the negative stigma against chemistry which states that chemistry is dangerous and difficult (Ardura & Pérez-Bitrián, 2018; Kartikawati & Azizah, 2017). This difference in perception affects students' perspective in learning chemistry, which is known as the concept of learning. Conception is an individual's response to something based on a different learning experience. Differences in experience during learning can lead to different perspectives on student learning so that the understanding they have is also different (Ananda & Fadhilaturrahmi, 2018; Mansur & Rafiudin, 2020; Wahyuningtyas & Sulasmono, 2020). The conception of learning refers to students' views of their personal experiences and learning contexts (Koopman et al., 2011; O'Keefe et al., 2021). Research on the conception of learning was initiated and recognized as the initial foundation by Saljo in 1979 by interviewing 90 students about the meaning of learning for them (Lee et al., 2008; Tsai et al., 2011). The conception of student learning can be classified into two types of conception, namely reproductive and constructivist conceptions. This means that learners with reproductive conceptions perceive simply learning science as knowledge used to achieve higher test scores. On the other hand, students with constructivist conceptions believe that studying science makes them better understand science concepts to get a more valuable meaning in life. Tsai's theory was then updated by categorizing the conception of learning into low and high categories with six aspects (Lee et al., 2008; Sobri et al., 2020). Low-level conceptions include memorizing, testing, counting, and practicing. On the other hand, higher-order conceptions include increasing knowledge, applying it, understanding it, and seeing it in new ways. A study of high school students in Indonesia in science subjects found that students dominantly owned the low-level conception of learning. The aspect of remembering was considered the basis that needed to be possessed to master the higher aspects (Rachmatullah et al., 2018). This habit needs to be corrected by bringing the concept of learning to a higher level and not only limited to learning science. This is because learning is a hierarchical system; the higher a person's conception of learning, the greater the tendency to have an in-depth approach to learning (Marton et al., 1993; Zheng et al., 2018).

The Rasch model is considered an effective, reliable, and modern approach to assessing the validity and reliability of scales used in various scientific fields (Sari et al., 2016; Spinou et al., 2019). Rasch model is the most commonly applied model concerning the relationship between item difficulty and respondent's ability (Spinou et al., 2019). Analysis using the Rasch model has been widely used to determine the psychometric properties of measuring scales of instruments such as the Self-Efficacy Teaching and Knowledge Instrument for Science Teachers, quality of Teacher Success questionnaire, Evaluating the Quality of Teaching for Students' Creativity, and many other relevant studies that indicate the use of the Rasch model is a trending study among researchers (Bui et al., 2020; Pruski et al., 2017; Tabatabaee-Yazdi et al., 2018). Previous research has investigated students' conceptions of science learning (Lin et al., 2015; Sadi, 2015; Sadi & Çevik, 2016; Sadi & Lee, 2017). Other research findings also found that there are differences in students' conceptions of learning science in different subjects, so the measurements must be carried out specifically in each subject, such as chemistry, biology, and physics (Chiou et al., 2012; Lee et al., 2008). If we want to understand students' conceptions of biology learning, the instrument must only focus on biology (Sadi & Lee, 2017). Chemistry subjects that are considered difficult and abstract need more attention to map the learning conceptions possessed by students (Refriwati, 2015; Zammiluni et al., 2018). This aims to prepare appropriate methods and strategies for learning, both remotely and face-to-face, to improve the learning conception of secondary school students to a higher level. Therefore, there is a need for research to develop learning conception instruments that specifically measure students' chemistry learning conceptions during learning during the pandemic that is valid and ready to be used. There is no research study on learning conception instruments that specifically measure the conception of chemistry learning. This research aims to develop a learning conception instrument that specifically measures the conception of chemistry learning. It is hoped that this instrument can assist teachers in measuring the conception of chemistry learning in students.

## 2. METHODS

This research is part of development research that focuses on determining the quality of the instruments developed before use. Qualitative and quantitative methods were used to determine the quality. The qualitative method in question is the synthesis of instruments by modifying the instrument as a form of development of the instruments that have been developed (Soltani & Askarizadeh, 2021; Zheng et al., 2018). An instrument in the form of a questionnaire was developed specifically to measure students' learning conceptions in learning chemistry during learning during a pandemic. The questionnaire consists of 24 items using five categories of the Likert scale (strongly disagree, disagree, neutral, agree, strongly agree) with six dimensions or aspects, namely memorizing (KP29, KP30, KP31, & KP32), testing (KP33, KP34, KP35, & KP36), counting and practicing (KP37, KP38, KP39, & KP40), increasing knowledge (KP41, KP42, KP43, & KP44), applying (KP45, KP46, KP47, & KP48), and understanding knowledge and seeing chemistry in a new way (KP49, KP50, KP51, & KP52). Furthermore, the qualitative method was also applied to validate the instrument construct based on consultation with two Yogyakarta University lecturers regarding the instrument that had been developed. This validation is carried out by involving validators who are experts in their fields to obtain both scientific and content instruments (Widyaningsih & Yusuf, 2018). Meanwhile, quantitative methods were applied to the analysis of the psychometric quality of the instrument to determine to construct validity and item reliability using Rasch analysis. The psychometric qualities of the questionnaire measured included reliability, item suitability statistics (Infit & Outfit values and Point-Measure Correlation Coefficient (PTMEA Corr), the functionality of the scale category, and item bias (Differential Item Functioning (DIF)) were analyzed using the Winsteps application version 3.73 (Linacre, 2009).

The research data was collected using an online survey via a google form. Researchers spread an online questionnaire link through chemistry subject teachers at schools, then distributed it to students. Students fill out questionnaires voluntarily without coercion from researchers or subject teachers. 247 high school students in Padang City for the 2021/2022 academic year participated in this study to see their perspective on the developed learning conception instrument. They consisted of 78 males and 169 females from various classes, namely ten, eleven, and twelve, with the distribution according to Table 1. The technique used to analyze the data was descriptive qualitative, and quantitative analysis.

**Table 1.** Demographic profile of respondents

	Category	Frequency	Percentage (%)
School	State	181	73
	Private	66	27
Gender	Man	78	32
	Woman	169	68
Class	X	145	59
	XI	20	8
	XII	82	33

## 3. RESULTS AND DISCUSSION

### Results

At the development stage, a chemistry learning conception questionnaire consisting of 24 items (KP29 – KP52) was generated with positive statements. Based on input and corrections from experts/experts, the questionnaire was revised by replacing six with negative statements. The questionnaire totaled 18 positive statements and six negative statements (KP31, KP35, KP40, KP44, KP47, and KP50). For example, item KP31 is "I learn chemistry not by memorizing what the teacher teaches." The constructively valid questionnaire was based on expert/expert correction, then empirical validity was tested using Rasch analysis to determine the quality of the instrument. Based on the results of data analysis, it is known that the average value of the logit person and item is positive, namely the logit person (respondent) value of 0.77 with a standard deviation of 0.73. The respondent has an average ability above the average item difficulty. Meanwhile, the item has a logit value of 0.00 and a standard deviation of 0.42, indicating that the item has a standard level of difficulty. Person reliability (0.83) is in the good category (good), and item reliability (0.96) is in the excellent category. In the Rasch model, the reliability value ranges from 0 to 1. If the value is below 0.5, it means that it can be said to be unreliable or unacceptable reliability; 0.5-0.67 (poor); 0.67-0.80 (enough); 0.81-0.90 (good); 0.91-0.94 (very good); and above 0.94 (excellent) (Fisher, 2007). The reliability results can also be seen from the Cronbach alpha value (0.88) with the good category. These results indicate similarities between the two methods of measuring

reliability in the Rasch analysis, which generally describes the interaction between the person and the items of the conception of chemistry learning. As indicated by the separation score, the grouping of persons and items has a score of 2.24 with a good category for the person and 5.00 for items categorized as very good. This score indicates that the items on the instrument can very well separate or group respondents based on the level of conception of learning chemistry. Meanwhile, the score of separation or respondents showed that the respondents involved in this study were heterogeneous from five groups with different levels of conception. The results of the next analysis are item suitability statistics that provide an overview of the level of difficulty (measure or logit score), the level of measurement accuracy (Standard Error Measurement score), and the suitability of each item in the chemistry learning conception questionnaire. The results show that Item KP35 (logit score 1.12) is the most difficult item to approve by respondents, and KP49 (logit score -0.19) is the most easily approved item. Most items have an accuracy level of 0.08 (except for five items, namely KP35, KP44, KP31, KP50, and KP47, with a score of 0.07).

For item suitability (fit or misfit of an item), the Infit and Outfit Mean Square (MNSQ) scores are the most important statistical indicators to identify the model's fit accurately. Acceptable MNSQ scores range from 0.5 to 1.5 (Meyer, 2014). The results of data analysis showed that 4 of the 24 items, namely KP31, KP35, KP44, and KP50, were in the unacceptable score range with the MNSQ infit score (1.72; 2.31; 2.04; and 1.54), respectively. and MNSQ outfits (1.86; 2.51; 2.21; and 1.57), respectively. One other item is also not in the acceptance range of the MNSQ outfit score, namely KP47 (1.55), although the MNSQ infit score is in the acceptable range (1.47). Another indicator that states the suitability of an item is the Point Measure Correlation Coefficient by estimates the polarity of the item. Items were categorized as very good (score 0.4-0.8), good (score 0.30-0.39), moderate (0.20-0.29), and failed in item polarity (score <0.19 ) (Rosli et al., 2020). The results show that the KP31, KP35, and KP44 are unacceptable, with successive scores of 0.17, -0.24, and 0.09. Meanwhile, the KP49 item is the item with the highest polarity (0.77) but is still in the acceptable range (very good). Based on the analysis, five items (KP31, KP35, KP44, KP47, and KP50) must be eliminated from the instrument because they do not meet the acceptance threshold score for item suitability statistics (infit and outfit MNSQ and Point Measure Correlation Coefficient as discussed previously. The item (KP47) does not meet the MNSQ outfit acceptance score, and the other four items (KP31, KP35, KP 44, and KP50) do not meet the MNSQ infit and outfit acceptance score and the item polarity score (Bond et al., 2020). It indicates that these items cannot measure the same as other items properly. The next analysis is the analysis of the functioning of the questionnaire with a categorical scale, whose results are shown in Table 1.

**Table 2.** The functionality of categories on a questionnaire with a Likert scale

Category	Data Calculation Category used	Frequ ency (%)	Average of Observed Measurements	Infit MNSQ	Outfit MNSQ	Andrich-Threshold Measurement	Category Measurement
1: Strongly Disagree	224	4	0.18	1.67	1.93	None	-2.50
2: Disagree	424	7	0.14	1.14	1.16	-0.85	-1.21
3: Neutral	1958	33	0.27	0.73	0.69	-1.29	-0.12
4: Agree	1871	32	0.91	0.86	0.79	0.73	1.16
5: Strongly Agree	1451	24	1.53	0.90	0.92	1.41	2.78

Categories 5, 4, and 3 are easily approved by most respondents, with the percentages being 24%, 32%, and 33%, respectively. Meanwhile, categories 2 and 1 were only filled by a small number of respondents, namely 7% and 4%. The MNSQ infit and outfit scores as an indicator of the functioning of the category scale for the four categories (5, 4, 3, and 2) were in the acceptable range. Meanwhile, category 1 (strongly disagree) on the scale has an MNSQ infit score (1.67) and an MNSQ outfit (1.93) exceeds the acceptance limit, with an ideal score of 1.0 and a score above 1.5 being considered problematic (Linacre, 2009). In the observed measurement mean score column, the score did not increase consistently from categories 1 to 5, namely from 0.18; 0.14; 0.27; 0.91; and 1.53. The existence of an inverted score (up and down in categories 1 and 2) indicates that respondents are confused in agreeing with the statements from the two categories. This is in line with the results in the Andrich-Threshold measurement score column. The Andrich-Threshold measurement score represents the probability points between the two previous

adjacent categories (2 and 3, 3 and 4, and so on) on a category scale. Therefore, the first category has an unknown score (none). The score obtained shows the non-ideality of the Andrich-Threshold measurement score. The increase in the score is not consistent with the increase in categories on the category scale (with scores respectively -0.85; -1.29; 0.73; and 1.41). The illustration supports this in Figure 1, where all categories have peaks and are separated from each other on the curve, although category 2 has a relatively low peak.

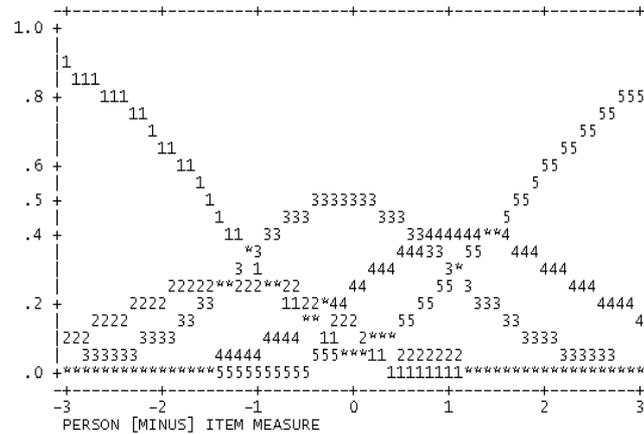


Figure 1. The categories probability curve of the chemistry learning conception questionnaire

Based on the measurement results, the category scale has functioned well to separate respondents based on their level of learning conception. This score can be interpreted that respondents with a low conception of chemistry learning (-2.50) will tend to choose category 1 (Strongly Disagree). On the other hand, respondents with a high conception of chemistry learning (2.78) will choose category 5 (Strongly Agree). For more details, the relationship between the difficulty level of the item to be approved and the person describing the respondent's level of chemistry learning conception can be seen in Figure 2. Figure 2 shows a wright map between item items and person, describing the relationship between item difficulty level to be approved and respondent's ability. Item number KP35 with a sequential logit value of 1.12 is the item that has the highest level of difficulty or is difficult to approve. However, these items have not been able to measure students with a conceptual ability to learn chemistry above it (from 180P, 210P, etc.) to students 054P (logit score 2.86), who are students with the highest level of conception of learning chemistry. In addition, item KP33 (with a logit score of -0.39) is classified as the item with the lowest difficulty level or the easiest to approve. However, these items have not been able to reach four students with a lower conception of chemistry, namely 158P (logit score: -0.72), 166P (logit score: -0.72), 240P (logit score: -0.82), and 089P (logit score -1.57) which is the student with the lowest level of chemistry learning conception.

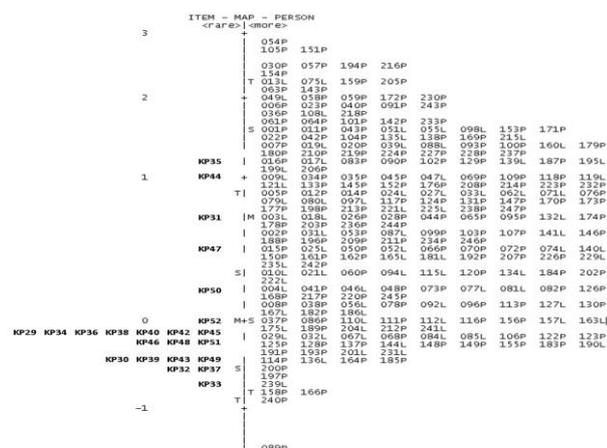


Figure 2. Wright map or distribution map of the relationship between item difficulty and a person's ability

The last analysis is the item refractive index (differential item functioning (DIF) with the aim of testing whether there are differences in responses from individual subgroups with a certain level of similarity, such as gender (male and female) that can have an impact on the measurement results. DIF is an additional aspect of compatibility with the Rasch model, which may affect validity by comparing data between subgroups through a biased scale score. DIF can be assessed by comparing item response functions between groups of people in the sample on a measured construct. In this study, DIF analysis was carried out by grouping respondents based on gender, namely male and female high school students, to find out whether there is a bias in the items of students' learning conception instruments based on the perceptions of male and female students. The results of the refractive index analysis can be seen in the DIF measurement score column, which states the level of difficulty in each group. Furthermore, the difference in DIF is obtained on the same item, which is one indicator of item bias of an item. The item's refractive index category includes large (score >0.64 logit), moderate (score range 0.43-0.64), and acceptable (score <0.43 logit) (Zwick, 2012). The analysis results show that item KP40 is classified as biased in the medium category. Based on data analysis, it is known that four items, namely KP35, KP36, KP40, and KP50, indicated a bias towards male and female respondents.

## Discussion

Based on the study results, it is known that the instrument developed is valid based on the expert correction. However, it needs to be revised by replacing some positive items with negative items. The combination of positive and negative items on an instrument is intended to reduce misunderstandings in measurement because of the tendency of someone to agree to a statement without understanding its content quickly or filling it with a certain response pattern (Desstya et al., 2019; Schlimbach & Asghari, 2020; Zeng et al., 2020). In addition, the item contributes to the validity of the measurement by broadening the way individuals think and organizing their beliefs about the construct under study (Rosana et al., 2017; Samsudin et al., 2021; Weijters & Baumgartner, 2012). All items, including six revised statements, were declared valid because they measured the desired attributes, namely six aspects of the concept of learning chemistry developed, namely memorizing (KP29, KP30, KP31, & KP32), testing (KP33, KP34, KP35, & KP36), calculating and practice (KP37, KP38, KP39, & KP40), increase knowledge (KP41, KP42, KP43, & KP44), apply (KP45, KP46, KP47, & KP48), and understand knowledge and see chemistry in a new way (KP49, KP50, KP51, & KP52) (Soltani & Askarizadeh, 2021; Zheng et al., 2018). Judging from the suitability of the items, 5 out of 24 items had to be eliminated because the measurement data did not match the Rasch model and were declared invalid. The instrument can be said to be valid when it represents the fit of the measurement data to the fit model from Rasch (Jang & Protacio, 2020; Rosli et al., 2020; Segers et al., 2018). Five items (KP31, KP35, KP44, KP47, and KP50) do not meet the acceptance threshold value for item conformity statistics and must be eliminated because these items illustrate the difficulty of respondents in understanding and agreeing with the items (Juditya et al., 2020; Wilmskoetter et al., 2019). Of the five eliminated items, four (KP31, KP44, KP47, and KP50) were items with negative statements. For example, KP47 reads, "I study chemistry not to be applied in everyday life." The use of a survey instrument consisting of a mixture of positive and negative statements can affect the validity and reliability of the instrument (Chan & Ismail, 2014; Chyung et al., 2018). It may be due to the tendency of respondents to fill in inconsistently when the instrument consists of negative and positive items, and respondents are less careful in understanding the difference in meaning between the two types of statement items (Colosi, 2005; Roszkowski & Soven, 2010; Sonderen et al., 2013).

The results of research related to the functioning of the Likert scale of the instrument prove that the four categories of the scale can be accepted or function well (categories 2, 3, 4, and 5) in determining respondents' approval based on the conception of chemistry learning, except for category 1 (strongly disagree), so it is necessary considered combining the two categories (1 and 2) into four scales only. The reduction of these categories is intended to obtain accurate and optimal measurement results (Irwanto et al., 2017; Vanzile-Tamsen, 2017). Furthermore, the research results related to the index show that four items, namely KP35, KP36, KP40, and KP50, are biased towards male and female respondents. Of the four items, two of them were eliminated (KP35 and KP 50) because they did not meet the acceptance of the item suitability value. The other two items (KP36 and KP40) are still used by first making revisions to avoid the possibility of different respondents' answers based on certain genders (Boone et al., 2014). These two items are negative statement items that are converted into positive statements in the hope of not confusing students, both male, and female, to determine and choose a category scale from this statement. For example, item KP36 which initially reads "I don't think studying chemistry has a close relationship with chemistry exams," was revised into a positive statement, "Learning chemistry has a close relationship with exams." This is because teachers always conduct evaluations in the form of daily tests, mid-term exams, or semester exams on all subjects at school, including chemistry subjects. This statement

is considered biased because students in different groups (male and female) have Differences in agreeing items are caused by the confusion of students to determine the choice of approval category, which is considered far deviated from the perceived reality so that these negative statements tend to form different dimensions (Merritt, 2012). In addition, learners with low reading or cognitive skills failed to notice changes in item wording or failed to adjust their responses (Bolt et al., 2020; Steedle et al., 2019; Steinmann et al., 2021).

#### 4. CONCLUSION

From the overall results of the analysis, the instrument is considered feasible to be tested because it is valid in content and is empirically and reliable. These results indicate similarities between the two methods of measuring reliability in the Rasch analysis, which generally describes the interaction between the person and the items of the conception of chemistry learning. The 24 instrument items were concluded by the requirements of the Rasch model and could be used to measure the chemistry learning conception of high school students.

#### 5. REFERENCES

- Ananda, R., & Fadhilaturrehmi, F. (2018). Analisis Kemampuan Guru Sekolah Dasar dalam Implementasi Pembelajaran Tematik di SD - Neliti. Jurnal Basicedu. <https://doi.org/10.31004/basicedu.v2i2.138>.
- Ardura, D., & Pérez-Bitrián, A. (2018). The effect of motivation on the choice of chemistry in secondary schools: adaptation and validation of the science motivation questionnaire ii to spanish students. *Chemistry Education Research and Practice*, 19(3), 905–918. <https://doi.org/10.1039/C8RP00098K>.
- Bolt, D., Wang, Y. C., Meyer, R. H., & Pier, L. (2020). An IRT mixture model for rating scale confusion associated with negatively worded items in measures of social-emotional learning. *Applied Measurement in Education*, 33(2), 331–348. <https://doi.org/10.1080/08957347.2020.1789140>.
- Bond, T. G., Yan, Z., & Heene, M. (2020). *Applying the Rasch Model* (4th ed.). Routledge.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer Netherlands.
- Bui, T. L. T., Kazarenkov, V. I., & Tran, V. D. (2020). Application of rasch model to develop a questionnaire for evaluating the quality of teaching for students' creativity development. *International Journal of Learning, Teaching and Educational Research*, 19(8), 278–296. <https://doi.org/10.26803/ijlter.19.8.15>.
- Chan, S. W., & Ismail, Z. (2014). Developing Statistical Reasoning Assessment Instrument for High School Students in Descriptive Statistics. *Procedia - Social and Behavioral Sciences*, 116. <https://doi.org/10.1016/j.sbspro.2014.01.943>.
- Chiou, G. L., Liang, J. C., & Tsai, C. C. (2012). Undergraduate students' conceptions of and approaches to learning in biology: a study of their structural models and gender differences. *International Journal of Science Education*, 34(2), 167–195. <https://doi.org/10.1080/09500693.2011.558131>.
- Churiyah, M., Sholikhah, S., Filianti, F., & Sakdiyyah, D. A. (2020). Indonesia education readiness conducting distance learning in covid-19 pandemic situation. . . *International Journal of Multicultural and Multireligious Understanding*, 7(6). <https://doi.org/10.18415/ijmmu.v7i6.1833>.
- Chyung, S. ., Barkin, J. R., & Shamsy, J. A. (2018). Evidence-based survey design: the use of negatively worded items in surveys. *Performance Improvement*, 57(3), 16–25. <https://doi.org/10.1002/pfi.21749>.
- Colosi, R. (2005). Negatively worded questions cause respondent confusion. *Proceedings of the Survey Research Methods Section*, 35(1), 113–130. <https://doi.org/10.1.1.556.243>.
- Desstya, A., Prasetyo, Z. K., Suyanta, Susila, I., & Irwanto. (2019). Developing an instrument to detect science misconception of an elementary school teacher. *International Journal of Instruction*, 12(3), 201–218. <https://doi.org/10.29333/iji.2019.12313a>.
- Fadhilaturrehmi, Ananda, R., & Yolanda, S. (2021). Persepsi Guru Sekolah Dasar terhadap Pembelajaran Jarak Jauh di Masa Pandemi Covid 19. *Jurnal Basicedu*, 5(3), 1683–1688. <https://doi.org/10.31004/basicedu.v5i3.1187>.
- Fisher, W. P. J. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, 21(1).
- Irwanto, Rohaeti, E., Lfx, E. W., & Suyanta. (2017). Development of an Integrated Assessment Instrument for Measuring Analytical Thinking and Science Process Skills. *AIP Conference Proceedings*, 1847(050005), 1–6. <https://doi.org/10.1063/1.4983907>.
- Jang, B. G., & Protacio, M. S. (2020). Use of Cognitive Interviews to Test the Validity of a Reading

- Motivation Instrument for English Learners. *Reading Psychology*, 1–21. <https://doi.org/10.1080/02702711.2020.1813226>.
- Juditya, S., Suherman, A., Ma'mun, A., & Rusdiana, A. (2020). The Basic Movement Skill Test Instrument of Ball Games for Students Aged 13-15 Years. *Jurnal Pendidikan Jasmani Dan Olahraga*, 5(1). <https://doi.org/10.17509/jpjo.v5i1.21447>.
- Kartikawati, A., & Azizah, U. (2017). Keterampilan Proses Sains Peserta Didik Melalui Penerapan Model Pembelajaran Learning Cycle 7-E pada Materi Laju Reaksi Kelas Xi di SMA Negeri 1 Krebun. *Unesa Journal of Chemical Education*, 6(2), 229–237.
- Koopman, M., Brok, P. Den, Beijaard, D., & Teune, P. (2011). Learning processes of students in pre-vocational secondary education: Relations between goal orientations, information processing strategies and development of conceptual knowledge. *Learning and Individual Differences*, 21(4). <https://doi.org/10.1016/j.lindif.2011.01.004>.
- Lee, M. H., Johanson, R. E., & Tsai, C. C. (2008). Exploring Taiwanese high school students' conceptions of and approaches to learning science through a structural equation modeling analysis. *Science Education Journal*, 92(2), 191–220. <https://doi.org/10.1002/sce.20245>.
- Lin, T. C., Liang, J. C., & Tsai, C. C. (2015). Conceptions of memorizing and understanding in learning, and self-efcacy held by university biology majors. *International Journal of Science Education*, 37(3), 446–468.
- Linacre, J. M. (2009). *A user's guide to winsteps*. Winsteps.
- Mansur, H., & Rafiudin, R. (2020). Pengembangan Media Pembelajaran Infografis untuk Meningkatkan Minat Belajar Mahasiswa. *Jurnal Komunikasi Pendidikan*. <https://doi.org/10.32585/jkp.v4i1.443>.
- Marton, F. M., Dall' Alba, G., & Beaty, E. (1993). Conceptions of learning. *International Journal of Educational Research*, 19(3), 277–300.
- Merritt, S. . (2012). The two-factor solution to allen and meyer's (1990) affective commitment scale: effects of negatively worded items. *Journal of Business and Psychology*, 27(4), 421–436. <https://doi.org/10.1007/s10869-011-9252-3>.
- Meyer, J. (2014). *Applied measurement with jmetrik* (1st ed.). Routledge.
- O'Keefe, P. A., Lee, H. Y., & Chen, P. (2021). Changing Students' Beliefs About Learning Can Unveil Their Potential. *Policy Insights from the Behavioral and Brain Sciences*, 8(1), 84–91. <https://doi.org/10.1177/2372732220984173>.
- Pruski, L. A., Blanco, S. L., Riggs, R. A., Grimes, K. K., Fordtran, C. W., Barbola, C. M., Cornell, J. E., & Lichtenstein, M. J. (2017). Construct validation of the self-efficacy teaching and knowledge instrument for science teachers-revised (SETAKIST-R): lessons learned. *Journal of Science Teacher Education*, 24(7), 1133–1156. <https://doi.org/10.1007/s10972-013-9351-2>.
- Rachmatullah, A., Diana, S., & Ha, M. (2018). Identifying Indonesian upper-secondary school students' orientations to learn science and gender effect through the use of structural equation modeling. *Journal of Baltic Science Education*, 17(4), 663–648. <https://doi.org/10.33225/jbse/18.17.633>.
- Refriwati. (2015). Improvement students' learning motivation using problem-based learning approach on chemistry topic of tsm grade 11 SMKN 1 bukit sundi Kabupaten Solok. *Jurnal Pendidikan Indonesia*, 1(1), 36–42. <https://doi.org/10.29210/1201525>.
- Rosana, D., Ramadani, M., & Dewi, K. (2017). The Instrument for Assessing the Performance of Science Process Skills Based on Nature of Science (NOS). *Jurnal Cakrawala Pendidikan*, 36(3), 435–445. <https://doi.org/10.21831/cp.v36i3.14731>.
- Rosli, R., Abdullah, M., Siregar, N. C., Shazana, N., Hamid, A., Abdullah, S., Halim, L., Daud, N. M., Bahari, S. A., Majid, R. A., & Bais, B. (2020). Student awareness of space science : rasch model analysis for validity and reliability. *World Journal of Education*, 10(3), 170–177. <https://doi.org/10.5430/wje.v10n3p170>.
- Roszkowski, M. ., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, 35(1), 113–130. <https://doi.org/10.1080/02602930802618344>.
- Sa'diyah, K. (2021). Pengembangan E-Modul Berbasis Digital Flipbook Untuk Mempermudah Pembelajaran Jarak Jauh Di SMA. *Edukatif: Jurnal Ilmu Pendidikan*, 3(4), 1298–1308. <https://doi.org/10.31004/edukatif.v3i4.561>.
- Sadi, Ö. (2015). The analysis of high school students' conceptions of learning in different domains. *International Journal of Environmental and Science Education*, 10(6), 813–827. <https://doi.org/10.12973/ijese.2015.278a>.
- Sadi, Ö., & Çevik, M. (2016). Investigating of conceptions of learning biology with respect to gender, grade level and school type. In *SHS Web of Conferences (Vol. 26)*. <https://doi.org/10.1051/shsconf/20162601025>.

- Sadi, Ö., & Lee, M. H. (2017). Exploring Taiwanese and Turkish high school students' conceptions of learning biology. *Journal of Biological Education*. <https://doi.org/10.1080/00219266.2017.1285799>.
- Samsudin, A., Bias Cahyani, P., Rusdiana, D., Aminudin, A. H., Coştu, B., Samsudin, A., Cahyani, P. B., Rusdiana, D., & Coştu, A. A. H. (2021). Development of a multitier open-ended work and energy instrument (MOWEI) using Rasch analysis to identify students' misconceptions. *Cypriot Journal of Educational Sciences*, 16(1), 16. <https://doi.org/10.18844/cjes.v16i1.5504>.
- Sari, D. R., Sekarwana, N., Hinduan, Z. R., & Sumintono, B. (2016). Analisis tingkat kepuasan masyarakat terhadap dimensi kualitas pelayanan tenaga pelaksana eliminasi menggunakan pemodelan rasch. *Jurnal Sistem Kesehatan*, 2(1), 47–55. <https://doi.org/10.24198/jsk.v2i1.10419>.
- Schlimbach, R., & Asghari, R. (2020). Das Digital Canvas: Ein Instrument zur Konzeption digitaler Geschäftsmodelle. *HMD Praxis Der Wirtschaftsinformatik*, 57(4), 866–878. <https://doi.org/10.1365/s40702-020-00624-9>.
- Segers, M., Martens, R., & Bossche, P. Van den. (2018). Understanding how a case-based assessment instrument influences student teachers' learning approaches. *Teaching and Teacher Education*, 4(3). <https://doi.org/10.1016/j.tate.2008.02.022>.
- Sobri, M., Nursaptini, N., & Novitasari, S. (2020). Mewujudkan Kemandirian Belajar Melalui Pembelajaran Berbasis Daring Diperguruan Tinggi Pada Era Industri 4.0. *Jurnal Pendidikan Glasser*, 4(1), 64. <https://doi.org/10.32529/glasser.v4i1.373>.
- Soltani, A., & Askarizadeh, G. (2021). How students' conceptions of learning science are related to their motivational beliefs and self-regulation. *Learning and Motivation*, 73. <https://doi.org/10.1016/j.lmot.2021.101707>.
- Sonderen, V., Sanderman, R., & Coyne, J. . (2013). Ineffectiveness of reverse wording of questionnaire items: let's learn from cows in the rain. *PloS One*, 8(7). <https://doi.org/10.1371/journal.pone.0068967>.
- Spinou, A., Siegert, R. J., Guan, W., Patel, A. S., Gosker, H. R., Lee, K. K., Elston, C., Loebinger, M. R., Wilson, R., Garrod, R., & Biring, S. S. (2019). The development and validation of the bronchiectasis health questionnaire. *European Respiratory Journal*, 49(5). <https://doi.org/10.1183/13993003.01532-2016>.
- Steedle, J. T., Hong, M., & Cheng, Y. (2019). The effects of inattentive responding on construct validity evidence when measuring social-emotional learning competencies. *Educational Measurement: Issues and Practice*, 38(2), 101–111. <https://doi.org/10.1111/emip.12256>.
- Steinmann, I., Strietholt, R., & Braeken, J. (2021). A constrained factor mixture analysis model for consistent and inconsistent respondents to mixed-worded scales. *Psychological Methods Advance Online Publication*. <https://doi.org/10.1037/met0000392>.
- Tabatabaee-Yazdi, M., Motallebzadeh, K., Ashraf, H., & Baghei, P. (2018). Development and validation of a teacher success questionnaire using the rasch model. *International Journal of Instruction*, 11(2), 129–144. <https://doi.org/10.12973/iji.2018.11210a>.
- Tsai, C. C., Ho, H. N. J., Liang, J. C., & Lin, H. M. (2011). Scientific epistemic beliefs, conceptions of learning science and self-efficacy of learning science among high school students. *Learning and Instruction*, 21, 757–769. <https://doi.org/10.1016/j.learninstruc.2011.05.002>.
- Vanzile-Tamsen, C. (2017). Using rasch analysis to inform rating scale development. *Research in Higher Education*, 58(8), 922–933. <https://doi.org/10.1007/s1162-017-9448-0>.
- Wahyuningtyas, R., & Sulasmono, B. S. (2020). Pentingnya Media dalam Pembelajaran Guna Meningkatkan Hasil Belajar di Sekolah Dasar. *Edukatif: Jurnal Ilmu Pendidikan*, 2(1), 23–27. <https://doi.org/10.31004/edukatif.v2i1.77>.
- Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, 49(5), 737–747. <https://doi.org/10.1509/jmr.11.0368>.
- Widyaningsih, S. W., & Yusuf, I. (2018). Analisis soal modul Laboratorium Fisika Sekolah I menggunakan rasch model. *Gravity: Jurnal Ilmiah Penelitian Dan Pembelajaran Fisika*, 4(1), 33–46. <https://doi.org/10.30870/gravity.v4i1.3116>.
- Wijayanti, R. M., & Fauziah, P. Y. (2020). Perspektif dan Peran Orangtua dalam Program PJJ Masa Pandemi Covid-19 di PAUD. *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini*, 5(3), 1304–1312. <https://doi.org/10.31004/obsesi.v5i2.768>.
- Wilmskoetter, J., Bonilha, H., Hong, I., Hazelwood, R. J., Martin-Harris, B., & Velozo, C. (2019). Construct validity of the Eating Assessment Tool (EAT-10). *Disability and Rehabilitation*, 41(5), 549–559. <https://doi.org/10.1080/09638288.2017.1398787>.
- Zammiluni, Ulianas, A., & Mawardi. (2018). Development of guided inquiry based work sheet with class and laboratory activity on chemical bonding topic in senior high school. *International Journal of*

- Chemistry Education Research*, 2(1). <https://doi.org/10.20885/ijcer.vol2.iss2.art1>.
- Zeng, B., Wen, H., & Junjie, Z. (2020). How does the valence of wording affect features of a scale? the method effects in the undergraduate learning burnout scale. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.585179>.
- Zheng, L., Dong, Y., Huang, R., Chang, C. Y., & Bhagat, K. K. (2018). Investigating the interrelationships among COLS, ALS, and self-efficacy in learning science. *International Journal of Science Education*, 40(1), 139–158. <https://doi.org/10.1080/09500693.2017.1402142>.
- Zwick, R. (2012). *A review of ets differential item functioning assessment procedures: flagging rules, minimum sample size requirements, and criterion refinement*. <https://doi.org/10.1002/j.2333-8504.2012.tb02290.x>.